Plundering Coalitions*

Weijia Li† Alireza Mohammadi Sepahvand‡

April 9, 2026

Abstract

We develop a model to study coalitions that extract the resources of outsiders. The players in our model are endowed with power and resources. The ruling coalition plunders outsiders, distributes the plundered resources among its members, and guarantees that insiders' resources remain safe. Our analysis focuses on the resilience of the equilibrium ruling coalition to exogenous shocks affecting the power and resources of both insiders and outsiders, as well as the intensity of plundering. We show that a coalition with a classical hierarchical structure—where power and resources are equal within each "rank" but strictly higher in higher ranks—produces greater resilience to external shocks affecting outsiders' power and resources. The only exception arises when plundering intensity is relatively weak, in which case the internal distribution of power and resources does not affect external resilience. Our final results provide insights into how the intensity of plundering impacts the internal and external resilience of ruling coalitions across political environments.

Keywords: Political Economy, Coalition Formation, Institutions, Resilience, Plundering.

1 Introduction

Coalition formation is always challenging (Ray and Vohra (2015a)), and a "plundering coalition" is no exception. For such a coalition, the wealth they can distribute among

*This paper is extracted from the doctoral dissertation of Alireza Mohammadi Sepahvand, completed as part of his Ph.D. studies at Monash University. We thank Arthur Campbell, Yves Zenou, Matthew Jackson, John Moore, Kaveh Majlesi, Massimo Morelli, Paul Raschky, Vai-Lam Mui, Sam Jindani, Solmaz Moslehi, Melika Liporace, Pawel Gola, and seminar and conference participants at Monash University, and the University of Edinburgh for helpful comments and suggestions. All errors are ours.

†Monash University, Australia. Email: Weijia.li@monash.edu.

‡Edinburgh University, UK. Email: mailto:amohamma@ed.ac.uk.

coalition members is plundered from the outsiders. This setup applies to a wide range of important social phenomena, such as an army that plunders the civil society, or an oligarchical government that extracts from its citizens (Puga and Trefler (2014); Xu (2018); Sánchez De La Sierra (2020); Henn et al. (2024)). We formally study the problem to form a coalition whose primary objective is to plunder outsiders. To our knowledge, this is the first such attempt in coalition formation games. Our model yields a series of novel results. Central to our model, we study the *resilience* of a plundering coalition against shocks to outsiders, which justifies the key organizational principle of hierarchy for an effective army or a stable oligarchy. We also propose a new methodology to analyze the resilience of an equilibrium coalition against exogenous shocks.

Our model features a society of finitely many individuals, each endowed with power and wealth. A coalition is "winning" if its aggregate *power* exceeds a supermajority threshold.[1] The game opens with an initial winning coalition, whose members sequentially propose alternative coalitions. A proposal is adopted—and the proposed coalition becomes the ruling coalition—if it commands a supermajority of power and is accepted by all of its members. If no proposal succeeds, the initial coalition rules by default. The ruling coalition then defeats the outsiders, plunders their wealth, and distributes the spoils among its members.

We are primarily interested in the properties of the ruling coalition. The ruling coalition is shaped by the following trade-off: admitting a new member raises the power of the coalition, which can plunder more wealth from defeated outsiders; but the new member is also costly because the coalition cannot plunder the wealth of the new member anymore. We show that the ruling coalition which optimally balances this trade-off exists, is unique, and admits a clean axiomatic characterization.

For the equilibrium ruling coalition, we then characterize a necessary and sufficient condition, which prepares our central analysis of coalitional resilience. The ruling coalition must outperform, in terms of plunder, two classes of alternatives: its own sub-coalitions, and any alliance between one of its sub-coalitions and a subset of outsiders. These two conditions motivate two distinct notions of resilience. A coalition is more internally resilient if it is more likely to survive exogenous shocks to the power and resources of its own members. It is more externally resilient if, holding its members' characteristics fixed, it is more likely to survive exogenous shocks to the power and resources of outsiders.

To understand the conditions of high external resilience, we conduct a thought experiment. Take any two coalition members and make them more homogeneous: transfer power from the stronger to the weaker, or wealth from the richer to the poorer, without

[1] Formally, a coalition is winning if its aggregate power exceeds β fraction of society's total power, with $\beta > 1/2$. Notice that here the super-majority refers to power, not to votes.

flipping their relative ranks. This transfer holds the characteristics of the ruling coalition constant, so it is still the unique ruling coalition. But importantly, such a transfer reduces the risk of the more threatening member with stronger power or lower wealth. After the transfer, the ruling coalition becomes more resilient to any alliance between a sub-coalition that includes the more threatening member and any subset of outsiders, where the outsiders are subject to any possible perturbation of their resources and power. At the same time, the ruling coalition is equally resilient to an alliance between a sub-coalition that includes the less threatening member and any subset of outsiders. Therefore, the ruling coalition becomes more externally resilient if two of its members become more homogeneous.

It is important to note that the analysis does not imply that absolute equality maximizes external resilience. Instead, the analysis implies that more externally resilient than others is a ruling coalition of a classic hierarchical structure. Such a hierarchical coalition is partitioned into well-defined "ranks." Within each "rank," all members are absolutely equal with each other; but higher "ranked" members are both richer and more powerful than lower ranked members. Once such a hierarchy emerges, it is not possible to further improve external resilience through an operation of transfer as above. Our analysis therefore offers a justification for the classical hierarchical structure of many organizations, such as armies and bureaucracy, by their unique capacity in bearing changes to its enemies/subjects. This justification is, as far as we know, novel, in contrast to the conventional emphasis on the advantage of a hierarchical structure in incentive-alignment (Qian (1994); Mookherjee (2013)) or division of labor in (Garicano (2000); Garicano and Rossi-Hansberg (2015)).

Finally, we study how internal and external resilience respond to changes in the plundering technology. A stronger plundering technology increases the cost of keeping a player inside the ruling coalition, since insiders' wealth is protected from extraction. This shifts members' preferences toward more exclusive alternatives—smaller, less powerful, and poorer sub-coalitions—and away from more inclusive ones. Because the internal threats to the ruling coalition are precisely these exclusive sub-coalitions, stronger plundering lowers internal resilience. In short, it makes the coalition harder to hold together.

For the external resilience of the ruling coalition, a stronger plundering technology is a double-edged sword. It makes exclusive alternatives involving small segments of society more threatening, but makes inclusive alternatives involving broader segments of society less threatening. As a result, the effect on external resilience depends on which alternatives are more likely to arise under external shocks. If shocks are more likely to generate exclusive alternatives, stronger plundering lowers external resilience. If shocks are more likely to generate inclusive alternatives, it raises external resilience. This implies that, in a power-light plundering environment, the ruling coalition may be more resilient to a stronger and wealthier opposition than to a weaker and poorer one.

Lastly, although the effect of plundering intensity on external resilience generally depends on the internal configuration of power and resources within the ruling coalition, we identify a broad class of political environments in which this is not the case. In these environments, plundering is power-intensive—for example, because property rights are better protected. In particular, for every exclusive alternative, there exists a more threatening inclusive alternative. As a result, external resilience depends only on how plundering intensity affects the attractiveness of inclusive alternatives. Since stronger plundering makes inclusive alternatives less attractive, it always increases external resilience in such environments. Thus, in power-intensive plundering environments, stronger plundering creates a trade-off between external and internal resilience: it increases external resilience but lowers internal resilience, regardless of how shocks are distributed. This implies that even imperfect property rights—which do not fully prevent plundering by insiders—may hinder the ruling coalition from achieving both internal and external stability when plundering technology changes. By contrast, in power-light plundering environments, a change in plundering intensity may alleviate both internal and external threats to the ruling coalition.

1.1 Relevant Literature

Our paper is relevant to a few strands of literature. The literature on coalition formation largely focus on characterizing the equilibrium coalition (Acemoglu et al. (2008); Ray and Vohra (2015b); Battaglini (2021)), or define stability mainly by incorporating the notion of "farsightedness" (Harsanyi (1974); Ray and Vohra (2015c)). We instead take one step further by analyzing the resilience of the equilibrium coalition against exogenous shocks. By doing so, we make a methodological contribution by proposing a simple framework to analyze the resilience of the equilibrium coalition. This novel focus on resilience also uncovers a lot of new substantive insights.

We bring together the two strands of literature on coalition formation and organizational economics of hierarchy. Existing literature usually focus on how a hierarchy may improve incentive-alignment or division of labor (Qian (1994); Qian et al. (2006); Mookherjee (2013); Garicano (2000); Garicano and Rossi-Hansberg (2015)). We offer a new justification for hierarchy: we show that a hierarchy is uniquely resistant to arbitrary exogenous changes to the characteristics of individuals outside the hierarchy. Our novel justification might be relevant to many hierarchies where the characteristics of outsiders are a first order concern, such as armies and fiscal bureaucracies (Besley and Persson (2009); Xu (2018); Sánchez De La Sierra (2020); Henn et al. (2024)).

Our model also makes novel contributions to a few central debates in political economy. First, political economists have uncovered that the interaction between power and wealth is a fundamental thread in political economy (Acemoglu and Robinson (2008); Dal Bó and Dal Bó (2011); Dal Bó et al. (2022); Acemoglu and Robinson (2013)). We contribute to this literature by an in-depth analysis of the power-wealth trade-off through the lens of coalition formation, the first ever attempt to our knowledge. It is through the coalition analysis that we uncover the innovative insight on the unique resilience of a hierarchical organization.

Our analysis also contributes to the burgeoning literature on political economy of non-democracies (Egorov and Sonin (2024)). Specifically, our analysis of internal and external resilience engages with the literature that addresses the trade-offs that authoritarian states resolve while dealing with internal or external threats to their rule. A strand of literature studies the loyalty-competence trade-off, i.e., how autocratic states balance the competence of their officials against their loyalty to prevent internal dissent (Besley and Kudamatsu (2007); Egorov and Sonin (2011); Jia et al. (2015); Zakharov (2016); Bai and Zhou (2019); Mattingly (2024)). Another strand of literature focuses on external problems such as mass protests, or propaganda (Wintrobe (1990); Wintrobe (2000); Konrad and Skaperdas (2007) Egorov et al. (2009); De Mesquita (2010); Yanagizawa-Drott (2014); Shadmehr (2018)). There are many trade-offs the dictators resolve while tackling external threats, for instance, the one between "informational openness" and "security" (Lorentzen (2013); Gehlbach and Sonin (2014); Lorentzen (2014); Guriev and Treisman (2019); Enikolopov et al. (2020)). Through the novel lens of coalition formation, we contribute to this literature by showing how the internal and external threats are related. In particular, we identify the condition for a trade-off between internal and external resilience driven by the process of coalition formation. Additionally, we provide insights into when this trade-off does not hold, and the characteristics of oppositions that can benefit an autocratic state engaging in intensive plundering.

The remainder of the paper is organized as follows. Section 2 introduces the model. Section 3 presents the preliminary analysis of the coalition formation game. Building on Section 3, we proceed by studying the resilience in Section 4. Section 5 concludes.

2 Environment

The society consists of a finite set of players $N = \{1, 2, \ldots, n\}$, and 2^N denote the set of all subsets of N. Time is finite and indexed by $t \in \{1, 2, \ldots, T\}$. The players are endowed with a pair of power p and resources x, specified by the mappings

$$p(\cdot) : N \to \mathbb{R}_{++},$$
$$x(\cdot) : N \to \mathbb{R}_{++}.$$

We refer to $p_i := p(i)$ and $x_i := x(i)$ as the political power and resources of individual $i \in N$. A coalition is any non-empty subset $I \subseteq N$, and each player belongs to at most one coalition at any stage of the game. The power and resources of any coalition $I \subseteq N$ are

$$P_I := \sum_{i \in I} p_i \quad \text{and} \quad X_I := \sum_{i \in I} x_i.$$

In particular, $P_N := \sum_{i \in N} p_i$ and $X_N := \sum_{i \in N} x_i$ are the power and resources of the society. A coalition I is a winning coalition if $P_I \geq \beta P_N$, where $\beta \in (1/2, 1]$ is a fixed supermajority requirement for power so that the winning coalition can defeat all outsiders. Note again that the supermajority requirement applies to power, not to votes. Denote the set of all winning coalitions by $\mathcal{W}$. There is a baseline payoff function $U : N \times \mathcal{W} \to \mathbb{R}$ that, for any player $i \in N$, assigns the payoff $U_i(I)$ if the winning coalition $I \in \mathcal{W}$ becomes the "ruling coalition." We also write $U(i, I) := U_i(I)$.

A ruling coalition of our model is necessarily a winning coalition. As the key novelty of our setup, a ruling coalition can only plunder outsiders, while the resources of its members are safe. This creates a central trade-off for our model. A new member who is brought into the ruling coalition strengthens its capability to plunder outsiders, but the ruling coalition loses the opportunity to plunder this new member. This key trade-off is formally captured by Assumption 1.

Assumption 1. *[Payoffs] For any $i \in N$ and $I \in \mathcal{W}$, $U_i(I) := x_i + w_i(I)$, where $w_i(\cdot)$ satisfies the following properties:*

1. *(Trade-off) If $I \in \mathcal{W} \setminus \{N\}$ and $i \in I$, we have $w_i(I) = G_i(P_I, X_I) > 0$, where $G_i(\cdot, \cdot) : [\beta P_N, P_N) \times [0, X_N) \to \mathbb{R}_{++}$ is a function continuous in P_I and X_I, satisfying the following conditions.*

   (a) For all I and $I' \in \mathcal{W} \setminus \{N\}$ with $P_I = P_{I'}$, if $i \in I$ and $i \in I'$, then $G_i(P_I, X_I) > G_i(P_{I'}, X_{I'})$ if and only if $X_I < X_{I'}$.

   (b) For all I and $I' \in \mathcal{W} \setminus \{N\}$ with $X_I = X_{I'}$, if $i \in I$ and $i \in I'$, then $G_i(P_I, X_I) > G_i(P_{I'}, X_{I'})$ if and only if $P_I > P_{I'}$.

2. *If $I \in \mathcal{W} \setminus \{N\}$ and $i \notin I$, then $w_i(I) < 0$.*

3. *For all $i \in N$, $w_i(N) = 0$.*

Assumption 1 establishes important primitives of the model. Part 1 introduces the key primitive, the function

$$G_i(\cdot,\cdot).$$

The function $G_i(\cdot,\cdot)$ ranks the plundered resources of any individual across non-trivial ruling coalitions of which she is a member.[2] Part 1(a) says that between ruling coalitions with equal power, players prefer the one with fewer internal resources, which permits more external resources for plundering. Meanwhile, between ruling coalitions with equal resources, players prefer the one with larger power (Part 1(b)), as it strengthens the ruling coalition in extracting resources. Both Part 1(a) and Part 1(b) imply that when the ruling coalition is not the grand coalition N, insiders obtain strictly positive payoffs from plundering outsiders. Together with Part 2, this implies that inclusion in the ruling coalition strictly benefits insiders relative to their initial resources, while exclusion strictly harms outsiders relative to their initial resources. Part 3 states that the players' payoff from the plundered resources is zero when the ruling coalition is N, since there are no outsiders to plunder.

Under Assumption 1, a ruling coalition is fully characterized by its power and resources. This keeps the model tractable by eliminating the complexities that arise when the specific combination of players inside the ruling coalition also matters. For the rest of this paper, we thus write $G_i(P_I, X_I)$ for player i's plunder gains in coalition I. Assumption 1 immediately yields the following lemma.

Lemma 1. *Under Assumption 1, any player $i \in N$ has strictly increasing and continuous indifference curves over (P, X). The variables P and X are the aggregate power and resources of ruling coalitions that include the player i; $(P, X) \in [\beta P_N, P_N) \times [0, X_N)$.*

The following assumption imposes essentially common preferences for players over coalitions, which simplifies notations throughout the paper. We later show that the main results continue to hold under a considerably weaker assumption.

Assumption 2. *For all $I \in \mathcal{W}$ and all $i \in I$, $G_i(P_I, X_I) := g(i)G(P_I, X_I)$, where $g(i) > 0$.*

Under Assumption 2, there are two components in a player's preference over ruling coalitions that include the player: an idiosyncratic component $g(i)$ and a common component $G(P_I, X_I)$, which depends on the aggregate powers and resources of the coalition, (P_I, X_I). This assumption implies that for all players, the indifference curves over

[2] For example, one can view $G_i(P_I, X_I)$ as a combination of a plundering component $F(I) : \mathcal{W} \to \mathbb{R}_{++}$ and a share component $\Pi(i, I) : N \times \mathcal{W} \to [0, 1]$, i.e., $G_i(P_I, X_I) := \Pi(i, I)F(I)$ is the share allocated to individual i within the coalition I from plundered resources $F(I)$.

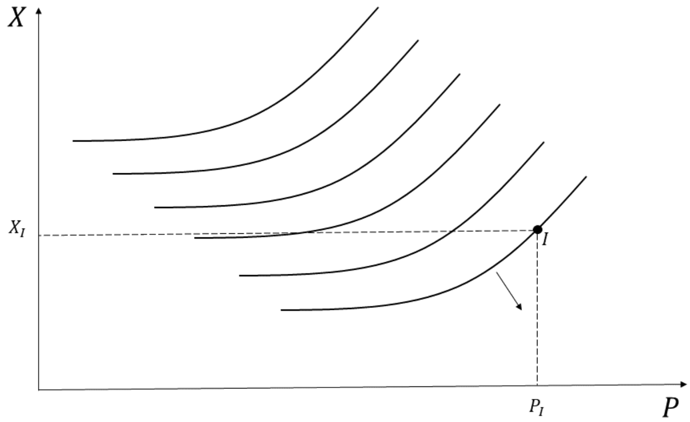

Figure 1: Identical indifference curves under Assumption 2

the coalitions containing them are the same and determined by the function $G(\cdot)$ (Figure 1). In other words, for any ruling coalitions $I, I' \in \mathcal{W}$ and any $i, j \in I \cap I'$, we have $U_i(I) \geq U_i(I')$ if and only if $U_j(I) \geq U_j(I')$, i.e., the preferences of players over any pair of ruling coalitions including them are identical. We slightly abuse the notation and denote:

$$G(I) := G(P_I, X_I),$$

The function $G(I)$ is particularly useful to characterize equilibrium coalition and its resilience.

Discussion on Assumption 2. Appendix B microfounds Assumption 2. It holds, for instance, when insiders' payoffs decompose as $w_i(I) = \Pi_i(I)F(I)$, where $\Pi_i(I)$ is an intra-coalition share (e.g., proportional to p_i/P_I) and $F(I)$ is the coalition's total extractable surplus, depending only on aggregate characteristics $(P_I, X_N - X_I)$.

Assumption 2 is also not essential. All results carry through under a weaker condition requiring preference consistency only on the set of "potential ruling coalitions" Z, defined later in Definition 3.1.[3] This weaker condition is also natural: since we focus on the survival of the ruling coalition under external shocks, it is reasonable to require that members weakly prefer to be in that coalition before any shock occurs—otherwise external resilience is trivially zero.

Definition 1. *Fix a function $G(\cdot, \cdot)$ that satisfies Assumption 1. For any ruling coalition I, denote the indifference curve through I by $X := H_I(P)$, which is implicitly defined by*

[3] As we will see, the proofs invoke preference comparisons only among coalitions in $\mathcal{Z}$ and the sub-coalitions and deviation coalitions generated by Z that enter the resilience constraints.

$G(P, X) = G(P_I, X_I)$.

Throughout the paper, we assume that the joint power and resources mapping is generic in the sense that for all $I, I' \in \mathcal{W}$, we have $P_I \neq P_{I'}$ or $X_I \neq X_{I'}$.[4] The following assumption helps establish the uniqueness results in the subsequent section.

Assumption 3. *Fix the power and resource mappings. Then, for all $I, I' \in \mathcal{W}$, we have $G(I) \neq G(I')$.*[5]

This assumption implies that players receive strictly different payoffs from different ruling coalitions involving them.

3 Preliminary Analysis of the Coalition Formation Game

This section establishes existence and uniqueness of the coalitional equilibrium, which prepares our analysis of its "resilience," i.e., how the equilibrium responds to exogenous shocks to players' power or resources.

3.1 Axiomatic analysis

We begin with an axiomatic analysis. As in Acemoglu et al. (2008) and Acemoglu et al. (2012), our axiomatic analysis shows that our results are independent of the details of the agenda-setting and voting protocols in the non-cooperative game introduced in Section 3.2. The axiomatic analysis will also help characterize the equilibrium of the non-cooperative game in Section 3.2.

Define a correspondence $\phi : \mathcal{W} \rightrightarrows 2^N$, which identifies the set of ruling coalitions corresponding to each initial winning coalition. We assume that ϕ satisfies the following axioms:

Axiom 1 (Non-triviality)**.** *For any $I \in \mathcal{W}$, $\emptyset \notin \phi(I)$ and $N \notin \phi(I)$.*

Axiom 2 (Super-majority of Power)**.** *For any $I \in \mathcal{W}$ and any $I' \in \phi(I)$, we have $I' \in \mathcal{W}$.*

Axiom 3 (Rationality)**.** *For any $I \in \mathcal{W}$, any $I' \in \phi(I)$, and any $I'' \in \mathcal{W}$,*

$$I'' \notin \phi(I) \iff G(I'') < G(I').$$

[4] Mathematically, this assumption is without much loss of generality, since the set of vectors $\{(P_I, X_I)\} \in \mathbb{R}_{++}^{2^{|N|+1}}$ that are not generic is the union of finitely many hyperplanes and therefore has Lebesgue measure zero.

[5] This assumption is also made without much loss of generality, as the set of functions from $\mathbb{R}^2$ to $\mathbb{R}$ for which the outputs coincide on a finite set of distinct inputs forms a measure-zero set in the space of all functions from $\mathbb{R}^2$ to $\mathbb{R}$.

These axioms are natural and capture the economic forces that give rise to the subgame perfect equilibria of the game in Section 3.2. Axiom 1 requires ϕ to map any initial winning coalition to a non-trivial ruling coalition. Axiom 2 requires any ruling coalition selected by ϕ to be a winning coalition. Axiom 3 imposes payoff-based selection: if $I' \in \phi(I)$, then no winning coalition I'' with strictly lower $G(\cdot)$ can be selected, and conversely any winning coalition with strictly higher $G(\cdot)$ must be selected. Proposition 1 establishes that these axioms pin down a unique mapping under Assumptions 1–2, and that the correspondence is single-valued under Assumptions 1–3.

Proposition 1. 1. *(Existence) Under Assumptions 1–2, the unique correspondence that satisfies Axioms 1–3 is*

$$\phi(I) = \arg\max_{W \in \mathcal{W}} G(W).$$

2. *(Uniqueness) Under Assumptions 1–3, the correspondence ϕ is single-valued.*

Proposition 1 is straightforward. It shows that the ruling coalition is a winning coalition that maximizes plunder, i.e., it maximizes $G(W)$ among all $W \in \mathcal{W}$. Under Assumption 3, this coalition is unique.

3.2 The non-cooperative extensive game

We next define the extensive-form complete-information game

$$\Gamma = \big(N, I_0, p(\cdot), x(\cdot), \{U_i(\cdot)\}_{i \in N}, \beta\big),$$

where N is the set of players, I_0 is the initial winning coalition, $p(\cdot)$ and $x(\cdot)$ are the power and resource mappings, $\{U_i(\cdot)\}_{i \in N}$ are the payoff functions satisfying Assumption 1 and Assumption 2, and $\beta \in (1/2, 1]$ is the supermajority requirement. The game starts with the initial winning coalition $I_0 \in \mathcal{W}$, and the steps are as follows:

1. Nature randomly picks an agenda setter a_q from the initial winning coalition, with $q = 1$, where $q \in \{1, \ldots, |I_0|\}$ denotes the round of agenda setting and voting.

2. The agenda setter a_q proposes a coalition $I_q \subseteq N$. If $P_{I_q} < \beta P_N$, then the game proceeds to Step 4. Otherwise, Nature chooses an order of votes and the game proceeds to Step 3.

3. The voting process begins. The coalition I_q forms if and only if the proposal of a_q is accepted by *all* players in I_q. In this case, I_q becomes the ruling coalition, and

each player $i \in N$ receives payoff $U_i(I_q) = x_i + w_i(I_q)$. Otherwise, following the first rejection of the proposal, the game proceeds to Step 4.

4. If $q < |I_0|$, Nature randomly picks a *new* agenda setter $a_{q+1} \in I_0 \setminus \{a_1, a_2, \ldots, a_q\}$ and the game returns to Step 2. If $q = |I_0|$, then I_0 becomes the ruling coalition and each player $i \in N$ receives payoff $U_i(I_0) = x_i + w_i(I_0)$.

The solution concept is subgame perfect equilibrium (SPE). The extensive-form game specifies players' strategies in any such equilibrium. A pure strategy of any player $i \in I_0$ is a pair of functions $\sigma_i(h) = (v_i(h, \mathcal{P}), \mathcal{P}_i(h))$ specifying her behavior at each decision node h: the function $v_i(h, \mathcal{P})$ specifies player i's vote (either 'Yes' or 'No') in any history h where Nature selects her to vote on a proposal $\mathcal{P}$, and $\mathcal{P}_i(h)$ specifies the coalition that player $i \in I_0$ proposes if selected by Nature as the agenda setter in history h. According to the extensive-form game, if $i \in N \setminus I_0$, player i cannot propose a coalition throughout the game.[6] Thus, the strategy of any $i \in N \setminus I_0$ is the voting function $v_i(h, \mathcal{P})$, which assigns either 'Yes' or 'No' to any proposed ruling coalition $\mathcal{P}$ containing i in any history h where Nature selects her to vote on $\mathcal{P}$.

We now establish existence and uniqueness of the ruling coalition in the non-cooperative coalition formation game, a preliminary result that prepares our analysis of the equilibrium's resilience to exogenous shocks. We also show that the SPE outcome of the coalition formation game coincides with the ruling coalition characterized by the axiomatic approach in Section 3.1.

Proposition 2. 1. *(Existence) Suppose that Assumptions 1–2 hold and that $\phi(I_0)$ satisfies Axioms 1–3. Then, for any $I \in \phi(I_0)$, there exists a pure-strategy SPE σ_I that produces I as the ruling coalition. In this SPE, each player $i \in N$ receives payoff $U_i(I) = x_i + w_i(I)$.*

2. *(Uniqueness) Suppose that Assumptions 1–3 hold, that $\phi(I_0)$ satisfies Axioms 1–3, and that $\phi(I_0) = \{I\}$. Then, in any SPE, I is the ruling coalition. In particular, in any SPE, each player $i \in N$ receives payoff $U_i(I) = x_i + w_i(I)$.*

The intuition is straightforward given Assumptions 1–3 and the axiomatic characterization in Proposition 1, where $\phi(I_0) = \arg\max_{W \in \mathcal{W}} G(W)$ for any $I_0 \in \mathcal{W}$. Any ruling coalition I identified by the axiomatic analysis (i.e., $I \in \phi(I_0)$) can be supported by an SPE in which every agenda setter from I_0 proposes I, and every voter from I_0 accepts I and rejects any other proposal. Under the supermajority rule $\beta \in (1/2, 1]$, any coalition I' proposed before I must include at least one player from I. Since coalitions form under

[6] All results continue to hold if the game is modified so that all players can be both voters and proposers.

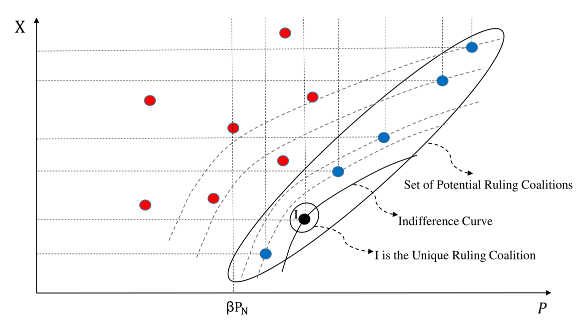

Figure 2: I is the unique ruling coalition of the game.

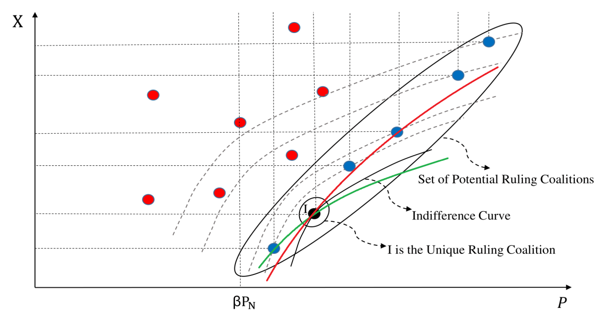

Figure 3: I is the unique ruling coalition under heterogeneous but consistent preferences over Z (each color represents one player's preferences over Z).

unanimity, the proposed strategy prevents any such I' from becoming the ruling coalition, ensuring that I forms. Moreover, in the axiomatic analysis, Assumption 3 implies that the correspondence $\phi : \mathcal{W} \to 2^N$ is single-valued, so any SPE yields the same ruling coalition.

Remark: potential ruling coalitions. We then define "potential ruling coalitions," which is a simple property of the equilibrium ruling coalition and a useful tool to understand coalitional resilience.

Definition 2. *For any power and resource mappings $p(\cdot)$ and $x(\cdot)$, define the set of potential ruling coalitions as*

$$Z := \{I \in \mathcal{W} \mid \nexists I' \in \mathcal{W} \text{ such that } P_{I'} > P_I \text{ and } X_{I'} < X_I\}. \tag{3.1}$$

Figure 2 illustrates the potential ruling coalitions (blue dots). A winning coalition is a potential ruling coalition if and only if no other winning coalition dominates it in both power and resources. The ruling coalition in Proposition 2 must therefore be a potential ruling coalition—a simple observation that will prove useful in what follows.

Remark: Robustness without Assumption 2. While Assumption 2 pins down a unique ruling coalition, our subsequent resilience analysis does not require it. We only need to assume a consistency of preferences on the set of potential ruling coalitions Z; in this case, a ruling coalition is already well defined (Figure 3) and we can proceed to characterize its resilience. As we will discuss below, analysis under Assumption 2 is actually informative on the general case with preference heterogeneity across members: preference heterogeneity tends to reduce external resilience, so results under Assumption 2 provide an upper bound on the external resilience of a ruling coalition whose members disagree about what constitutes a good alternative.

4 Main analysis on coalitional resilience

This section studies the resilience of the ruling coalition, which is the central part of the paper. Our analysis proceeds in three steps. We begin by characterizing the ruling coalition, which allows us to define "internal" and "external" resilience – robustness to changes in the power and resources of members versus outsiders (Proposition 3). The next two steps generate the central results of the paper. First, we show that the ruling coalition that is most externally resilient has a hierarchical structure (Proposition 4). Second, we uncover a potential trade-off between internal and external resilience, which depends on the "intensity" of power in plundering. We then comment on why our analysis of coalitional resilience may help understand the dynamics of the ruling coalition.

4.1 Internal and external resilience

This section shows that it is necessary and sufficient for the ruling coalition to dominate two types of threats: sub-coalitions of the ruling coalition and alternative coalitions that include players outside the ruling coalition. This distinction reflects the central challenges from regime insiders and outsiders (Svolik (2012); Meng (2020); Paine (2021); Egorov and Sonin (2024)), enabling us to conceptualize two notions of resilience. To proceed, we first define the set of "best sub-coalitions."

Definition 3. *For any $p(\cdot)$ and $x(\cdot)$ and any subset of players I, define the set of best sub-coalitions of I as follows:*

$$\mathcal{A}_I := \{A \subseteq I \,|\, A \neq \emptyset,\ \nexists A' \subseteq I \text{ such that } P_{A'} > P_A \text{ and } X_{A'} < X_A\}\,. \tag{4.1}$$

For any subset of players $I \subseteq N$, $\mathcal{A}_I$ includes the best subsets of I, i.e., those for which there does not exist another subset of I with both higher power and lower resources. Equation 4.1 is analogous to Equation 3.1 in Definition 3.1 for potential ruling coalitions, but restricts attention to sub-coalitions of the coalition in question. We can now prove Proposition 3, which characterizes the two types of threats to the ruling coalition. In particular, given a ruling coalition I, denote

$$A^{ins} \in (\mathcal{A}_I \backslash I) \cap \mathcal{W}$$

as a best sub-coalition of I that is also a winning coalition. A coalition A^{ins} is therefore an internal threat. Such a coalition A^{ins} can also form an alliance with outsiders to challenge the ruling coalition I, and these outsiders should also be best-subcoalitions of all outsiders. We denote a coalition of such outsiders as

$$A^{ext} \in \mathcal{A}_{N \setminus I}.$$

A ruling coalition only needs to defeat any internal threat A^{ins} and every alliance between any A^{ins} and most threatening outsiders A^{ext}. This intuition is established as follows.

Proposition 3. *Consider a game* $\Gamma = \left(I_0, p(\cdot), x(\cdot), \{U_i(\cdot)\}_{i \in N}, \beta\right)$ *and suppose that Assumptions 1–3 hold. Then* $\phi(I_0) = \{I\}$ *if and only if* $I \in \mathcal{W}$ *and:*

(i) For all $A^{\text{ins}} \in (\mathcal{A}_I \setminus \{I\}) \cap \mathcal{W}$, $G(I) > G(A^{\text{ins}})$ *(i.e., there is no profitable internal secession).*

(ii) For all $A^{\text{ext}} \in \mathcal{A}_{N \setminus I}$ *and for all* $A^{\text{ins}} \in \mathcal{A}_I$ *with* $A^{\text{ins}} \cup A^{\text{ext}} \in \mathcal{W}$, $G(I) > G(A^{\text{ins}} \cup A^{\text{ext}})$ *(i.e., there is no profitable external secession).*

Proposition 3 shows that a necessary and sufficient condition for I to defeat all alternative winning coalitions is to dominate (i) all its nontrivial best sub-coalitions and (ii) all alliances between its best sub-coalitions with best sub-coalitions of outsiders. The following example illustrates the proposition.

Motivated by Condition (i) of Proposition 3, we now define the key object for our analysis of internal resilience. We can express Condition (i) in the (P, X) space as follows, using the indifference curve $H_I(\cdot)$ over aggregate power P and aggregate resources X (Definition 1).

Definition 4. *For the ruling coalition* I*, the "internal safe area" is defined as*

$$\mathcal{S}_I^{\text{int}} := \left\{(P, X) \in \mathbb{R}_{++}^2 \;\middle|\; X > H_I(P)\right\}. \tag{4.2}$$

Then a ruling coalition I *has the same internal resilience as a ruling coalition* J *if and only if* $\mathcal{S}_I^{\text{int}} = \mathcal{S}_J^{\text{int}}$*. A ruling coalition* I *is strictly (weakly) more internally resilient than a ruling coalition* J *if and only if* $\mathcal{S}_J^{\text{int}} \subsetneq \mathcal{S}_I^{\text{int}}$ *(respectively,* $\mathcal{S}_J^{\text{int}} \subseteq \mathcal{S}_I^{\text{int}}$*).*

For any coalition I to be the ruling coalition, all its best sub-coalitions must lie in $\mathcal{S}^{\text{int}}$. This guarantees that all members of I prefer I to any best sub-coalition of I, satisfying Condition (i) of Proposition 3. We write $\mathcal{S}^{\text{int}}$ when there is no confusion. Figure 4 illustrates the internal safe area $\mathcal{S}^{\text{int}}$ for a coalition I. This set plays a central role in our subsequent analysis of internal resilience. In particular, we will see that a coalition I remains stable if, after an exchange of power and resources within I, all its best sub-coalitions remain inside the internal safe area $\mathcal{S}^{\text{int}}$.

We now turn to "external" threats. For any best sub-coalition of outsiders A^{ext} and any best sub-coalition of insiders A^{ins}, Condition (ii) of Proposition 3 requires $G(I) > G(A^{\text{ins}} \cup$

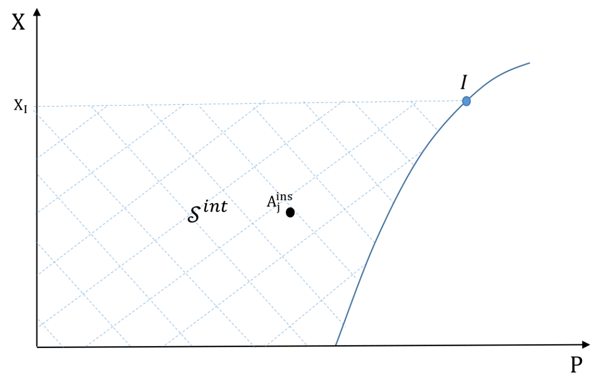

Figure 4: The internal safe area $\mathcal{S}^{\text{int}}$ is the shaded region.

A^{ext}), which is equivalent to $X_{A^{\text{ins}} \cup A^{\text{ext}}} > H_I(P_{A^{\text{ins}} \cup A^{\text{ext}}})$. Since aggregate power and resources are additive, this condition can be written as $X_{A^{\text{ins}}} + X_{A^{\text{ext}}} > H_I(P_{A^{\text{ins}}} + P_{A^{\text{ext}}})$, or

$$X_{A^{\text{ins}}} > H_I(P_{A^{\text{ins}}} + P_{A^{\text{ext}}}) - X_{A^{\text{ext}}}. \tag{4.3}$$

For any group of outsiders A^{ext} that satisfy Condition 4.3, the group of outsiders cannot threaten the ruling coalition. We need to make sure that Condition 4.3 is satisfied for every best sub-coalition of the ruling coalition. This motivates the following definition of the "external safe area" for a ruling coalition I.

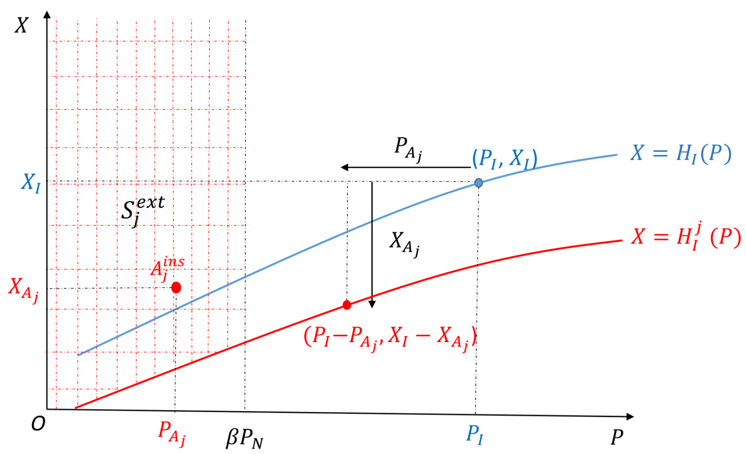

Figure 5: Construction of the shifted boundary $H_I^j(P) = H_I(P + P_{A_j}) - X_{A_j}$ (Definition 5). The red curve is obtained by translating $H_I(P)$ leftward by P_{A_j} and downward by X_{A_j} (L-shaped arrows). The shaded region $\mathcal{S}_j^{\text{ext}}$ collects all outsider coalitions that cannot form a profitable external deviation together with A_j^{ins}.

Definition 5. *Consider a ruling coalition I with its set of best sub-coalitions $\mathcal{A}_I$. For any $A_j^{\text{ins}} \in \mathcal{A}_I$, define*

$$H_I^j(P) := H_I(P + P_{A_j^{\text{ins}}}) - X_{A_j^{\text{ins}}}, \tag{4.4}$$

The function 4.4 is obtained by shifting the indifference curve $H_I(P)$ leftward by $P_{A_j^{\text{ins}}}$

units and downward by $X_{A_j^{\text{ins}}}$ *units.*

The area externally safe relative to A_j^{ins} *is*

$$\mathcal{S}_j^{\text{ext}} := \left\{(P, X) \in \mathbb{R}_{++}^2 \;\middle|\; P < \beta P_N \text{ and } X > H_I^j(P)\right\}. \tag{4.5}$$

Define the external safe area for the ruling coalition I *as*

$$\mathcal{S}_I^{\text{ext}} := \bigcap_{A_j^{\text{ins}} \in \mathcal{A}_I} \mathcal{S}_j^{\text{ext}}. \tag{4.6}$$

A ruling coalition I *has the same external resilience as a ruling coalition* J *if and only if* $\mathcal{S}_I^{\text{ext}} = \mathcal{S}_J^{\text{ext}}$. *A ruling coalition* I *is strictly (weakly) more externally resilient than a ruling coalition* J *if and only if* $\mathcal{S}_J^{\text{ext}} \subsetneq \mathcal{S}_I^{\text{ext}}$ *(respectively,* $\mathcal{S}_J^{\text{ext}} \subseteq \mathcal{S}_I^{\text{ext}}$*).*

What is the intuition behind Definition 5 and "external safety"? First, outsiders cannot have $P^{\text{ext}} \geq \beta P_N$. If they did, this outsider group would itself have supermajority power. Since $P_I \geq \beta P_N$ is required for I to be a ruling (therefore necessarily winning) coalition, such an outsider group can form a winning coalition on its own, and I could not remain the ruling coalition. Hence $P^{\text{ext}} < \beta P_N$ is a necessary condition for an outsider group to be "safe" for any ruling coalition I.

Second, fix a best sub-coalition of insiders $A_j^{\text{ins}} \in \mathcal{A}_I$. For any best sub-coalition of outsiders A^{ext} that lies exactly on the shifted indifference curve $H_I^j(P)$, insiders are indifferent between the current ruling coalition I and the alternative coalition $A_j^{\text{ins}} \cup A^{\text{ext}}$. If A^{ext} lies in the region $\mathcal{S}_j^{\text{ext}}$, then $A_j^{\text{ins}} \cup A^{\text{ext}}$ is strictly worse than I, so outsiders with such (P, X) cannot combine with A_j^{ins} to form a profitable external deviation. In this sense, $\mathcal{S}_j^{\text{ext}}$ is "externally safe" relative to A_j^{ins}. Taking the intersection over all $A_j^{\text{ins}} \in \mathcal{A}_I$ yields the external safe area $\mathcal{S}_I^{\text{ext}}$, i.e., the set of outsider best sub-coalitions that are simultaneously safe against *any* insider best sub-coalition.

Proposition 3 yields an additional insight: a ruling coalition must maintain a sufficiently high power-to-resource ratio relative to any relevant alternative, as captured by both the internal and external safe areas.[7] This, for instance, offers a rationale for the voluntary destruction of resources by a ruling coalition when faced with a threatening alternative.

Remark: Heterogeneity of preferences and resilience. How does preference heterogeneity within the ruling coalition affect external resilience once Assumption 2 is relaxed? Since the external safe area is defined as the region above the envelope of the

[7]Example 4 in the appendix demonstrates that neither the most powerful player nor the one with the fewest resources is necessarily included in the ruling coalition.

boundaries induced by insiders' best sub-coalitions, introducing an insider whose preferences differ from others can only add an additional boundary, which can only shift the envelope weakly upward. Hence, the external safe area can only weakly shrink, i.e., external resilience weakly decreases. The same logic applies to the internal safe area.

Now that we have provided a precise characterization of internal and external safe areas, we are ready to study internal and external resilience of a ruling coalition.

4.2 Which ruling coalitions are more externally resilient?

Consider a ruling coalition I and suppose that there are two players $i, j \in I$, with $p_i > p_j$ and $x_i < x_j$. Holding the power and resources of all other players fixed, transfer either (i) a portion of player i's power to player j, with $0 < \Delta p \leq \frac{p_i - p_j}{2}$, or (ii) a portion of player j's resources to player i, with $0 < \Delta x \leq \frac{x_j - x_i}{2}$ (Figure 6). The following proposition establishes that such an equalizing transfer between members (weakly) reduces the threat posed by the stronger or poorer member, and therefore (weakly) increases the coalition's external resilience. This result is general and does not depend on the precise specification of the plundering function $G(\cdot)$, i.e., on the shape of indifference curves.

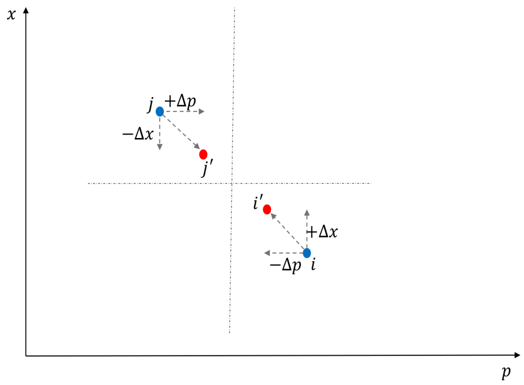

Figure 6: Exchange of powers and resources between player i and player j—from blue to red.

Proposition 4. *Suppose I is the unique ruling coalition of the game Γ, and there exist $i, j \in I$ with $p_i > p_j$ and $x_i < x_j$. Holding fixed the powers and resources of players in $I \setminus \{i, j\}$, consider the modified coalition $(I \setminus \{i, j\}) \cup \{i', j'\}$ where*

$$p_{i'} = p_i - \Delta p, \quad x_{i'} = x_i + \Delta x, \quad p_{j'} = p_j + \Delta p, \quad x_{j'} = x_j - \Delta x,$$

for any $0 < \Delta p \leq \frac{p_i - p_j}{2}$ and $0 < \Delta x \leq \frac{x_j - x_i}{2}$. Then the external resilience of $(I \setminus \{i, j\}) \cup \{i', j'\}$ is weakly higher than the external resilience of I.

Proposition 4 is the first key result on coalitional resilience. The proof is in Appendix A and proceeds in three steps.

Step 1. We identify two effects of the exchange from $\{i, j\}$ to $\{i', j'\}$: (i) best sub-coalitions of I that contain i but not j gain resources and lose power after the exchange, making them less threatening; and (ii) the set of best sub-coalitions may itself change—new ones may emerge and old ones may cease to qualify.

Step 2. We show that neither effect reduces external resilience. For effect (i), since the aggregate characteristics (P_I, X_I) are unchanged by the exchange, the indifference curve $H_I(\cdot)$ is unchanged, and the shifted boundary $H_A(P) := H_I(P + P_A) - X_A$ of any best sub-coalition A containing i but not j shifts strictly downward, weakly expanding the external safe area. For effect (ii), consider any post-exchange best sub-coalition A'. Its shifted boundary is weakly below the shifted boundary of its pre-exchange antecedent $A \subseteq I$—obtained by replacing i', j' with i, j—because A' has weakly lower power and weakly higher resources than A. Moreover, if $A \notin \mathcal{A}_I$, then some pre-exchange best sub-coalition $B \in \mathcal{A}_I$ dominates A, in the sense that $P_B > P_A$ and $X_B < X_A$. Since A' has weakly lower power and weakly higher resources than A, its shifted boundary lies weakly below that of A, and therefore also weakly below that of B. Hence the upper envelope of shifted boundaries cannot increase, and the external safe area cannot shrink.

Step 3. We show that these changes do not generate a profitable internal secession. Indeed, if a ruling coalition I already withstands internal secession before the exchange (Condition (i) of Proposition 3), any winning best sub-coalition after the exchange is weakly less threatening than some winning best sub-coalition before, which the original ruling coalition already defeated. Hence internal secession cannot be triggered by the exchange.

A new justification for hierarchy. As an important implication of Proposition 4, iterating the exchanges of power and resources depicted in Figure 6 yields "conditional equality" (or "conditional proportionality") within the ruling coalition. Coalition members are partitioned into ranks: within each rank, members have identical power and resources, and across ranks, power and resources are proportional, with the highest rank holding the most power and resources, followed by the second rank, and so on. At each step, external resilience weakly increases. Consequently, the resulting allocation has weakly higher external resilience than the initial allocation. In this sense, iterated exchanges produce a "hierarchy" (Figure 7). Our analysis thus offers a new perspective on why the most resilient plundering coalitions tend to exhibit hierarchical organization with well-defined ranks. Examples include stable oligarchies, Weberian bureaucracies, and armies.

It is important to note that comparing the external resilience of different "hierarchies"—i.e., different proportional configurations of power and resources within the ruling coalition—generally requires additional structure on the plundering function $G(\cdot)$. For instance, fix

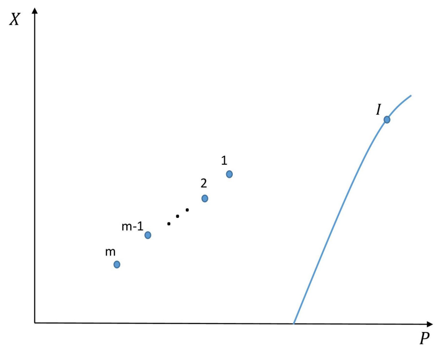

Figure 7: A coalition consisting of m ranks. Each blue dot represents a rank of players with identical power and resources and ranks are totally ordered by power P and resources X.

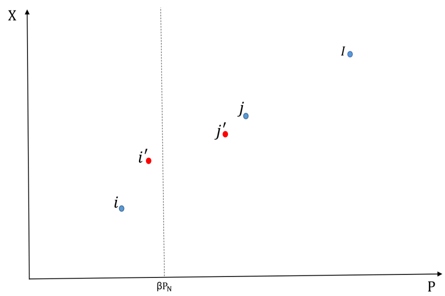

Figure 8: Ambiguity in comparing external resilience of proportional internal distributions.

the total power and resources of a two-player ruling coalition I and consider two internal configurations, $\{i, j\}$ and $\{i', j'\}$, such that $P_I = p_i + p_j = p_{i'} + p_{j'}$ and $X_I = x_i + x_j = x_{i'} + x_{j'}$ (Figure 8). Then there is no general argument that ranks which configuration yields higher external resilience without further restrictions on $G(\cdot)$ (e.g., beyond concavity of indifference curves).

Dynamic implications. It is natural to ask what happens after outsiders are plundered. Plundering depletes outsiders' resources and raises their power-to-resource ratios, potentially making them more attractive coalition members in the future and, in turn, destabilizing the very coalition that plundered them.

Our static game and resilience analysis can be interpreted as a benchmark for a repeated *stationary-bandit* environment: for each "season," the ruling coalition extracts and consumes the extracted resources, while outsiders rebuild resources through (agricultural) production before the next season. Each season then starts from a recovered distribution $x(\cdot)$ and the same coalition-formation and extraction problem is played again. In such a setting, extraction can persist period after period if the ruling coalition consume all resources and preserves outsiders' productive capacity (e.g., does not burn the land or kill the farmers), unless an exogenous shock disrupts the process. In this sense, our resilience analysis provides a simple benchmark for understanding how extractive coalitions evolve over time in the presence of shocks.

Preference heterogeneity and hierarchy. Importantly, the logic underlying Proposition 4 extends to heterogeneous preferences within the ruling coalition. We first set aside a degenerate case: if some insider strictly prefers a different ruling coalition, resilience is trivially zero under any external shock. Absent this, Proposition 4 extends under a weaker version of Assumption 2 that accommodates preference heterogeneity. The key force in the proof remains unchanged: the exchange shifts certain best sub-coalitions up and left in (P, X)-space, which pushes down the relevant boundary curves in Equation 4.4 that jointly defines the external safe area (Definition 5). Under Assumption 1, for any insider (regardless of the curvature of her indifference curves), a sub-coalition that moves up/left becomes weakly less attractive after the exchange. Hence, the associated boundary curve for that insider shifts weakly downward; aggregating across insiders, this weakly expands the external safe area. Therefore, the exchange in Figure 6 enlarges the external safe area in the same direction as in the benchmark case of homogeneous preferences under Assumption 2.

In Proposition 4, the exchange of power and resources *weakly* increases the external resilience of the ruling coalition. The next sections characterize conditions under which the exchange *strictly* increases external resilience, as well as conditions under which external resilience remains unchanged. In particular, we highlight the role of the shape of indifference curves—convex versus concave—which we interpret as capturing the strength of property-rights protection.

4.3 Increasing and decreasing return to power

Concave and convex indifference curves capture a key difference in how the marginal value of power varies with a coalition's power. For an indifference curve $H_I(P)$ that goes through a ruling coalition I, the marginal rate of substitution between power and resources is:

$$\frac{d}{dP} H_I(P).$$

This derivative measures the marginal value of power relative to resources. To see this, consider a small increase in a coalition's power by $\delta > 0$. To keep insiders indifferent, the coalition's internal resources needs to increase by approximately $\delta \cdot \frac{d}{dP} H_I(P)$. Recall that insiders dislike resources held *inside* the coalition, since these resources are protected from plunder. If $\frac{d}{dP} H_I(P)$ is large, a small increase in power can offset a large increase in internal resource that cannot be plundered, indicating a high marginal value of power. Based on the shapes of the indifference curves over (X, P)-space, we adopt the following definition.

Definition 6. *1. Under concave indifference curves ($\frac{d^2}{dP^2}H_I(P) < 0$), power displays decreasing marginal return relative to resources, or* ***decreasing return to power****.*

2. Under convex indifference curves ($\frac{d^2}{dP^2}H_I(P) > 0$), power displays increasing marginal return relative to resources, or ***increasing return to power****.*

Under concave indifference curves ($\frac{d^2}{dP^2}H_I(P) < 0$), the marginal value of power decreases as power increases. Consequently, a ruling coalition with many members, hence a high aggregate power P, has a low marginal value of power $\frac{d}{dP}H_I(P)$. This induces the ruling coalition to have relatively few members, so that the coalition attains an adequate level of power before its marginal value quickly declines, which also keeps enough resources outside and available to plundering.[8]

Under convex indifference curves, the marginal value of power increases as the coalition grows larger, so members are willing to accept the cost of additional insiders – namely, the resources shielded from plunder – in exchange for the gains from greater aggregate power. For instance, when plundering are constrained by institutions that protect property rights of everyone, these constraints might be neutralized only by a sufficiently powerful coalition.[9]

The following proposition shows that, under increasing returns to power, external resilience depends only on the coalition's aggregate power and resources, *not* on how these are distributed among its members.

Proposition 5 (Power-intensive plundering and invariance of external resilience)**.** *Suppose that preferences over coalitions (P, X) have strictly convex indifference curves. For any ruling coalition I and any $A \in \mathcal{A}_I \setminus \{I\}$, we have*

$$\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_A^{\text{ext}}.$$

Hence,

$$\mathcal{S}^{\text{ext}} := \bigcap_{A \in \mathcal{A}_I} \mathcal{S}_A^{\text{ext}} = \mathcal{S}_I^{\text{ext}}.$$

Therefore, any exchange of power and resources within I that preserves internal stability leaves the external resilience of I unchanged.

Proposition 5 is proved in Appendix A. Fix a ruling coalition I and an insider best sub-coalition $A^{\text{ins}} \in \mathcal{A}_I \setminus \{I\}$ that is also winning ($A^{\text{ins}} \in \mathcal{W}$). Under increasing return to

[8] This preference has an analogy in standard consumer theory: although a consumer prefers more of a good, diminishing marginal utility implies that extremely large amounts of the same good are not valuable at the margin.

[9] Example 3 in the appendix provides a more detailed discussion of environments that feature increasing return and decreasing return to power.

power, insiders strictly prefer larger ruling coalitions. Hence, for any best sub-coalition of outsiders $A^{\text{ext}} \in \mathcal{A}_{N\setminus I}$, insiders in A^{ins} prefer $A^{\text{ext}} \cup I$ to $A^{\text{ext}} \cup A^{\text{ins}}$. Any profitable and feasible external deviation of the form $A^{\text{ext}} \cup A^{\text{ins}}$ is therefore (weakly) dominated by the deviation $A^{\text{ext}} \cup I$. Thus, the set of outsider coalitions that can induce secession with A^{ins} is contained in the set that can induce secession with I, which implies that $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_1^{\text{ext}}$. Since I is itself included among the insider best sub-coalitions, we obtain

$$\mathcal{S}^{\text{ext}} := \bigcap_{A_i^{\text{ins}} \in \mathcal{A}_I} \mathcal{S}_i^{\text{ext}} = \mathcal{S}_I^{\text{ext}}.$$

Therefore, as long as internal exchanges of power and resources do not trigger internal secession, they leave $\mathcal{S}^{\text{ext}}$—and hence external resilience—unchanged.

Our next result, Proposition 6, shows that external resilience *strictly* increases as the ruling coalition becomes more hierarchical, provided that the plundering technology exhibits sufficiently decreasing return to power and Assumption 4 holds.

Assumption 4 (Winning high-power-to-resource bloc)**.** *For every mis-ordered pair $i, j \in I$ with $p_i > p_j$ and $x_i < x_j$, there exists $\lambda \in \left(\frac{p_j}{x_j}, \frac{p_i}{x_i}\right)$ such that $A_\lambda := \{k \in I : p_k - \lambda x_k > 0\} \in \mathcal{W}$.*

For any cutoff λ, the coalition $A_\lambda := \{k \in I : p_k - \lambda x_k > 0\}$ consists exactly of those insiders whose power-to-resource ratio exceeds the cutoff and, as we show in the Appendix, is *always* a best sub-coalition of I. Assumption 4 adds only that, for every mis-ordered pair $i, j \in I$ with $p_i > p_j$ and $x_i < x_j$, there exists a cutoff $\lambda \in \left(\frac{p_j}{x_j}, \frac{p_i}{x_i}\right)$ such that this best sub-coalition is winning. The assumption is relatively mild: it rules out only configurations in which insiders with relatively high power-to-resource ratios are, in the aggregate, too weak to form a winning coalition.

The assumption therefore implies that, for a relevant cutoff, A_λ is a *winning* best sub-coalition of I. This is crucial for external resilience, because only winning best sub-coalitions generate translated boundaries that enter the external safe area on $[0, \beta P_N)$. When an equalizing exchange lowers the power-to-resource ratio of a member of A_λ, the translated boundary generated by A_λ shifts strictly downward on $[0, \beta P_N)$. This downward shift is the key mechanism behind the strict gain from hierarchy in Proposition 6.

To measure concavity, we focus on a CES family of plundering functions $\{G_\rho\}_{\rho \neq 0}$. Fix a ruling coalition I with aggregate power and resources (P_I, X_I). For $\alpha \in (0,1)$ and $\rho \neq 0$, define

$$G_\rho(P, X) := \left[\alpha \left(\frac{P}{P_I}\right)^\rho + (1-\alpha)\left(\frac{\bar{X} - X}{\bar{X} - X_I}\right)^\rho\right]^{1/\rho},$$

where $\bar{X}$ is sufficiently large (with $\bar{X} > X_I$) and $(P, X) \in \mathbb{R}_+ \times (0, \bar{X})$. For $\rho \leq 1$, G_ρ is concave in (P, X); as ρ decreases, the marginal return to power decreases more rapidly (Figure 9).

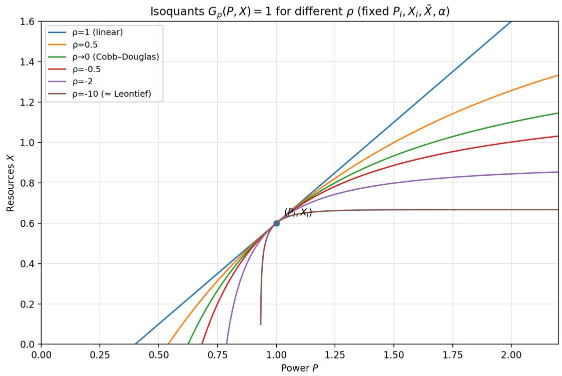

Figure 9: Isoquants of the CES plundering technology $G_\rho(P, X)$ for different values of ρ, normalized to pass through (P_I, X_I). Lower ρ implies a more concave/ Leontief-like shape.

Proposition 6 (Decreasing return to power and strict gains from hierarchy)**.** *Suppose that preferences are represented by $G_\rho(\cdot)$ for some $\rho < 1$, and that Assumption 4 holds. Let I be a ruling coalition, and let I' be the coalition obtained from I by iterating the bilateral exchanges in Figure 6 until the coalition reaches a hierarchical structure with well-defined ranks. Then there exists $\bar{\rho} < 1$ such that for all $\rho \leq \bar{\rho}$,*

$$\mathcal{S}_I^{\text{ext}} \subsetneq \mathcal{S}_{I'}^{\text{ext}}.$$

In other words, a shift to a hierarchy strictly increases external resilience when the marginal return to power is sufficiently decreasing.

The proof is in Appendix A.

Consider the sequence of bilateral exchanges within I described in Figure 6, repeated until the coalition becomes hierarchical. If the external safe area expands strictly at any intermediate step, then the final hierarchical coalition already has strictly greater external resilience, because Proposition 4 implies that each subsequent exchange weakly enlarges the external safe area. It is therefore enough to consider the last step of the sequence, at which there remains a mis-ordered pair $i, j \in I$ with $p_i > p_j$ and $x_i < x_j$.

The proof first constructs a best insider sub-coalition A^* with $i \in A^*$ and $j \notin A^*$. A sufficiently small exchange keeps the image of A^* winning and shifts its translated boundary strictly downward. Thus, to obtain a strict increase in the external safe area, it remains to show that, before the exchange, the boundary generated by A^* uniquely determines the external envelope on some nonempty interval.

To establish this, the proof considers the Leontief limit $\rho \to -\infty$, in which the indifference curve becomes an inverse-L with kink at (P_I, X_I) (Figure 10). In this limit,

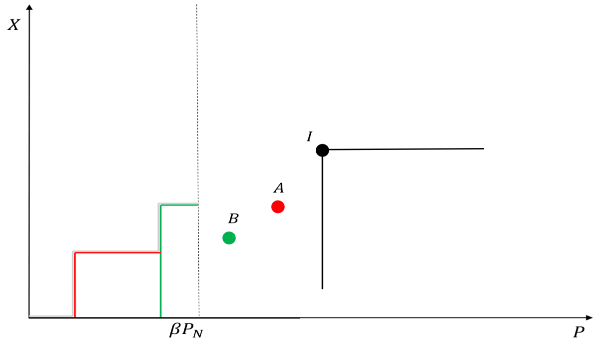

Figure 10: Under sufficiently concave (Leontief-like in limit) indifference curves, the boundary corresponding to a best sub-coalition is uniquely binding on the upper envelope that defines the external safe area.

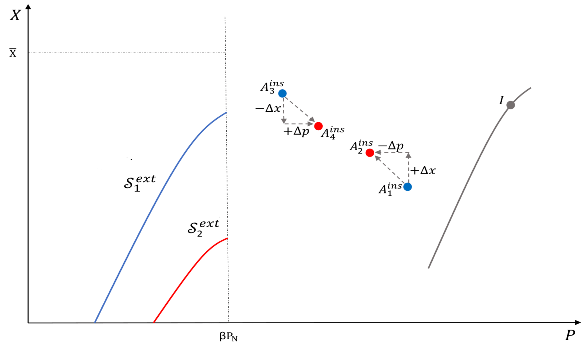

Figure 11: Under concave indifference curves, the exchange in Figure 6 expands the external safe area for best sub-coalitions that contain i (the member with higher power but lower resources than j).

the translated boundaries generated by winning best insider sub-coalitions become step functions, and their upper envelope is a staircase on $(0, \beta P_N)$. Assumption 4 implies that the coalition A^* constructed in the proof is also winning, so its boundary is part of this staircase. Hence the boundary generated by A^* uniquely determines the envelope on some nonempty open interval. Uniform convergence then implies that, for sufficiently small ρ, the same boundary continues to determine the CES envelope on a compact subinterval. Proposition 4 implies that all other relevant boundaries move weakly downward under the exchange, while the boundary generated by the image of A^* moves strictly downward. Therefore the external envelope falls strictly on that subinterval, so the external safe area expands strictly. Since all earlier exchanges weakly enlarge the external safe area, the transition to hierarchy strictly increases external resilience.

Weak property rights and the emergence of hierarchy. Our analysis identifies another central insight: the resilience advantage of a hierarchical coalition is stronger when property rights are sufficiently weak. Proposition 4 establishes that a move toward a hierarchical allocation never decreases external resilience for *any* plundering function satisfying Assumptions 1–3. Propositions 5 and 6 then sharpen this result by contrasting the two environments, interpreting the curvature of indifference curves as capturing the strength of property-rights protection. Under increasing return to power (strong property rights), a shift to a hierarchy leaves external resilience unchanged. Under sufficiently decreasing return to power (sufficiently weak property rights), and under Assumption 4, such a shift strictly increases external resilience. The organizational advantage of hierarchy therefore depends critically on the plundering environment: it is at most weakly beneficial in general, and becomes a strict gain when extraction is relatively unconstrained and the winning high-power-to-resource bloc condition holds.

4.4 Trade-off between internal and external resilience

We now link external and internal resilience by characterizing when a trade-off may arise between them.

4.4.1 The trade-off under increasing return to power

Under increasing return to power, a trade-off can arise between external and internal resilience as the plundering environment changes. In particular, if the marginal return to power further increases, this change lowers external resilience but raises internal resilience. To capture this intuition, we introduce a definition that allows us to uniformly compare the marginal return to power.

Definition 7. *The indifference curve $H_I(\cdot)$ displays a uniformly higher marginal return to power than $H'_I(\cdot)$ if and only if for all $P \in [\beta P_N, P_N]$,*

$$\frac{d}{dP}H_I(P) \;<\; \frac{d}{dP}H'_I(P),$$

denoted by $H \succ H'$.

The following proposition formalizes a fundamental trade-off between external and internal resilience under increasing return to power.

Proposition 7. *Fix a ruling coalition I. Suppose that the indifference curves $H_I(\cdot)$ and $H'_I(\cdot)$ display increasing return to power (i.e. convex). Then, if $H \succ H'$, the internal resilience of I under H is lower than under H', while its external resilience under H is higher than under H'.*

As established in Proposition 5, increasing return to power means that the most serious external deviation always involves outsiders paired with the *full* ruling coalition, rather than with a proper sub-coalition. This observation has a direct implication for how resilience responds to a shift in the plundering environment. As the marginal return to power increase uniformly, members place greater value on aggregate power, strengthening the appeal of larger coalitions. This works against external resilience: a stronger preference for larger coalitions makes it easier for outsiders to construct a profitable deviation with the full ruling coalition, shrinking the external safe area. But internal resilience moves in the opposite direction: precisely because aggregate power is more valuable, insiders are less tempted to defect to a smaller breakaway coalition that sacrifices power for a larger share of plunder. Increasing returns to power thus generates a fundamental trade-off: the same force that erodes external resilience strengthens internal cohesion.

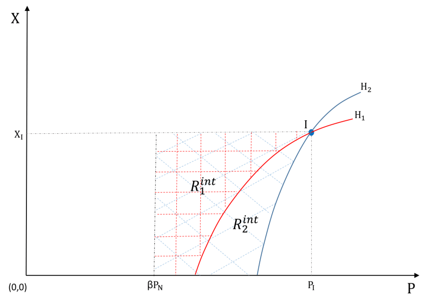

Figure 12: The increasing effect of a weaker plundering on internal resilience under concave indifference curves.

4.4.2 The trade-off under decreasing return to power

As in the convex case, a uniform increase in the marginal value of power increases internal resilience when indifference curves are concave (Figure 12). In this case, however, it is ambiguous how a change in marginal value of power affects external resilience, which now depends on the probability distribution of external shocks. The following example illustrates this point.

Example (the ambiguity of the trade-off under decreasing return to power) Consider a ruling coalition I and a plundering technology represented by the indifference curve H_1 passing through I. Suppose that the marginal value of power increases uniformly (e.g., stronger property rights), so that the relevant indifference curve moves from H_1 to H_2. Consider a best insider sub-coalition $A_1^{\text{ins}} \in \mathcal{A}_I$ that is realized with probability one. As Figure 13 illustrates, the effect of this shift on external resilience is ambiguous: depending on the distribution of external perturbations, the external safe area associated with A_1^{ins} may expand or shrink.

To see why, consider two distributions over external shocks, $\mathbb{P}_1^{\text{ext}}$ and $\mathbb{P}_2^{\text{ext}}$, as illustrated in Figure 13. Suppose $\mathbb{P}_1^{\text{ext}}$ places relatively more probability mass on the outsider sub-coalition A_2^{ext} than on A_1^{ext}, compared to $\mathbb{P}_2^{\text{ext}}$. Since internal and external shocks are independent, $A_2^{\text{ext}} \cup A_1^{\text{ins}}$ is more likely to arise than $A_1^{\text{ext}} \cup A_1^{\text{ins}}$ under $\mathbb{P}_1^{\text{ext}}$. Now suppose the shift from H_1 to H_2 renders $A_1^{\text{ext}} \cup A_1^{\text{ins}}$ newly preferred to I – a deviation that was not profitable under H_1. External resilience then falls under $\mathbb{P}_1^{\text{ext}}$, because the newly dangerous deviation receives relatively little probability weight under that distribution.

Conversely, suppose $\mathbb{P}_2^{\text{ext}}$ places relatively more mass on A_1^{ext} than on A_2^{ext}, so that $A_1^{\text{ext}} \cup A_1^{\text{ins}}$ is the more likely deviation under Pr_2^{ext}. Suppose instead that the shift from H_1 to H_2 renders $A_2^{\text{ext}} \cup A_1^{\text{ins}}$ no longer preferred to I, a deviation that *was* profitable under H_1. External resilience then rises under Pr_2^{ext}, because the deviation that becomes newly safe receives relatively little probability weight under that distribution.

In short, the effect of shifting from H_1 to H_2 on external resilience depends jointly

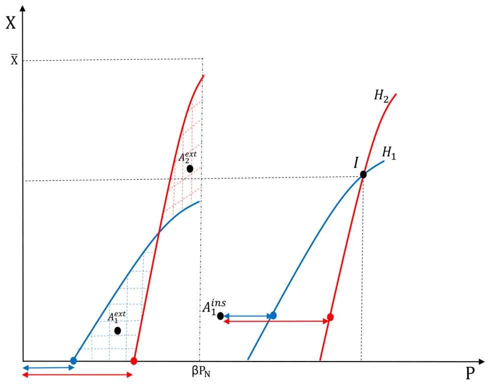

Figure 13: The ambiguous effect of a weaker plundering on external resilience.

on how the change in the indifference curve reclassifies outsider coalitions as dangerous or safe, and on where the external-shock distribution concentrates its mass. Notably, the trade-off with internal resilience need not hold: a shift from H_1 to H_2 can increase both internal and external resilience simultaneously. In this case, it might benefit the ruling coalition if it raise the protection of property rights for everyone, thereby uniformly increasing the marginal value of power. But notice that this process has a hard limit: when the plundering process reaches an increasing return to power, the trade-off between internal and external resilience must appear, so that further protection of property rights will unambiguously weaken external resilience.

4.5 Coalitional resilience: a summary

Table 1 summarizes our analysis of coalitional resilience. A shift to hierarchy never hurts external resilience for a plundering coalition; such a hierarchy actually displays the strongest advantage when properties are least protected. These results encode our central novel insight: hierarchy can emerge to fend off external threats to a plundering coalition, and it is especially likely to do so under weak protection of properties. Strong property rights further induces a stark trade-off between external and internal resilience, which is not necessarily the case under weak property rights.

5 Concluding remarks

This paper develops a theory of plundering coalitions. We conclude by commenting on further applications and directions for future work. Our theory shows that the ruling coalition has a relatively higher power-to-resources ratio than its alternatives. This result formalizes the concept of *Asabiyyah*, usually translated as "social cohesion," which is a central concept for understanding political dynamics in the Middle East (Khaldun (1967); Kuran (1996); Alatas (2014)). Specifically, the great historian and sociologist

	Increasing return to power *(strong property rights)*	**Decreasing return to power** *(weak property rights)*
General result	Hierarchy weakly increases external resilience.	
Effect of hierarchy on external resilience	Hierarchy leaves external resilience unchanged.	Hierarchy strictly increases external resilience, provided the return to power is sufficiently decreasing and Assumption 4 holds.
Trade-off between external and internal resilience	A trade-off arises between external and internal resilience.	No general trade-off arises between external and internal resilience.

Table 1: Hierarchy, the plundering environment, and coalitional resilience

Ibn Khaldun argued that nomadic tribes had a much higher level of "social cohesion" than urban civilizations, and that this strong social cohesion facilitated the conquest of urban civilizations by nomadic tribes. Our model microfounds the higher social cohesion of nomadic tribes through their relative poverty compared to urban civilizations, which generates a high power-to-resources ratio. It is therefore easier for nomadic tribes to form a coalition to plunder cities, which is a repeated pattern in the pre-modern world. Similar logic might also help understand communist revolutions (Morishima (1974); Roemer (1980); Roemer (1981); Brewer (2002)), where increasing inequality widens the power-to-resources gap and incentivizes the proletariat to rebel against the capitalists. Importantly, our analysis may provide a clue to understanding the oligarchic tendencies of these plundering coalitions. Our model may explain why successful nomadic conquerors and communist parties, even when starting as movements of radical equality, eventually evolved into strictly hierarchical structures.

In future research, our framework may be useful to study the resilience of coalitions to exogenous changes in players' characteristics and in the environment governing coalition formation. Moreover, although we study how coalitions respond to changes in power, resources, and plundering technology, these objects are exogenous in our model. Endogenizing them could be informative and suggests several extensions. A natural extension is to allow players to invest in power prior to coalition formation, which would clarify how the initial distribution of resources shapes power investment incentives and, ultimately, the ruling coalition. Another extension is to study environments in which the ruling coalition receives an exogenous flow of resources in addition to exploiting outsiders.

One could also consider endogenizing the plundering technology—interpreted as property-rights protection—in a dynamic version of our framework in which ruling coalitions can invest in institutions over time. This would contribute to the literature on the emergence and evolution of property-rights protection (Andolfatto (2002); Hafer (2006); Diermeier et al. (2017)) from the perspective of resilience. Moreover, institutions are persistent in many settings, so early institutional choices can have long-lasting effects on subsequent political and economic outcomes (Persson (2002); Michalopoulos and Papaioannou (2013); Lowes et al. (2017)). Finally, incorporating networks into the coalition-formation process is another promising direction. For example, König et al. (2017) studies how a network of military alliances affects conflict intensity. Extending our model along these lines could clarify how players' connections shape political alliances and their resilience.

Appendix A Proofs

Proof of Lemma 1. Fix $u \in \mathbb{R}$ and define the level set

$$I(u) := \{(P, X) \in [\beta P_N, P_N) \times [0, X_N) : G_i(P, X) = u\}.$$

We first show that G_i is strictly increasing in each argument. Fix X and let $P' > P$. By Assumption 1(1b), $G_i(P', X) > G_i(P, X)$. Fix P and let $X' < X$. By Assumption 1(1a), $G_i(P, X') > G_i(P, X)$. Hence, if $P' > P$ and $X' < X$, then $G_i(P', X') > G_i(P, X') > G_i(P, X)$.

We next show that the indifference curve is strictly increasing. Suppose, toward a contradiction, that there exist $(P_1, X_1), (P_2, X_2) \in I(u)$ with $P_2 > P_1$ and $X_2 \leq X_1$. If $X_2 = X_1$, then strict monotonicity in P gives $G_i(P_2, X_2) = G_i(P_2, X_1) > G_i(P_1, X_1)$, contradicting $G_i(P_2, X_2) = G_i(P_1, X_1) = u$. If $X_2 < X_1$, then by the implication above, $G_i(P_2, X_2) > G_i(P_1, X_1)$, again a contradiction. Therefore, along any indifference curve, a higher value of P must be accompanied by a higher value of X, so the indifference curve is strictly increasing.

Since G_i is strictly increasing in X, for each P there exists at most one X such that $(P, X) \in I(u)$. Thus, whenever $I(u)$ is nonempty, it is the graph of a function $X = H_i(P)$ on its domain.

It remains to show that H_i is continuous. Let $P_k \to P$ in the domain of H_i and set $X_k := H_i(P_k)$. Since $(P_k, X_k) \in I(u)$ for all k, any convergent subsequence $(P_{k_m}, X_{k_m}) \to (P, \bar{X})$ must satisfy, by continuity of G_i, $G_i(P, \bar{X}) = \lim_{m\to\infty} G_i(P_{k_m}, X_{k_m}) = u$. Hence $(P, \bar{X}) \in I(u)$. But $I(u)$ contains at most one point with first coordinate P, namely $(P, H_i(P))$, so $\bar{X} = H_i(P)$. Therefore every convergent subsequence of $\{X_k\}$ converges to $H_i(P)$, which implies $X_k \to H_i(P)$. Thus H_i is continuous.

Therefore, player i's indifference curves are strictly increasing and continuous. □

Proof of Proposition 1. Define, for each $I \in \mathcal{W}$, the correspondence $\phi(I) := \arg\max_{W \in \mathcal{W}} G(W)$. Since $\mathcal{W}$ is finite and non-empty, $\arg\max_{W \in \mathcal{W}} G(W)$ is well-defined and non-empty, so $\phi(I) \neq \emptyset$, establishing the first part of Axiom 1. Moreover, by construction, $\phi(I) \subseteq \mathcal{W}$, so Axiom 2 holds.

If $I' \in \phi(I)$, then $I' \in \arg\max_{W \in \mathcal{W}} G(W)$. Hence, for any $I'' \notin \phi(I)$, we must have $G(I'') < G(I')$. Conversely, if $G(I'') < G(I')$, then $I'' \notin \arg\max_{W \in \mathcal{W}} G(W)$, and therefore $I'' \notin \phi(I)$. This is exactly Axiom 3. Finally, Assumption 1 implies $G(N) = 0$, while for any $W \in \mathcal{W} \setminus \{N\}$ we have $G(W) > 0$. Hence N cannot be a maximizer of G over $\mathcal{W}$, so $N \notin \phi(I)$, establishing the second part of Axiom 1. This proves existence.

For uniqueness, suppose there is another correspondence ϕ' satisfying Axioms 1–3. Fix $I \in \mathcal{W}$ and take $I'' \in \phi'(I)$. If $I'' \notin \arg\max_{W \in \mathcal{W}} G(W)$, then let $I' \in \arg\max_{W \in \mathcal{W}} G(W)$. Since $G(I'') < G(I')$, Axiom 3 for ϕ' implies $I'' \notin \phi'(I)$, a contradiction. Hence $\phi'(I) \subseteq \arg\max_{W \in \mathcal{W}} G(W) = \phi(I)$.

Conversely, suppose $I' \in \phi(I)$ but $I' \notin \phi'(I)$. By Axiom 1, there exists $I'' \in \phi'(I)$. Since $I'' \in \phi'(I)$ and $I' \notin \phi'(I)$, Axiom 3 implies $G(I'') > G(I')$, contradicting $I' \in \phi(I) = \arg\max_{W \in \mathcal{W}} G(W)$. Therefore $\phi(I) \subseteq \phi'(I)$. Hence $\phi = \phi'$.

For the second statement, Assumption 3 implies that if $W, W' \in \mathcal{W}$ and $W \neq W'$, then $G(W) \neq G(W')$. Therefore the maximizer of G over $\mathcal{W}$ is unique, so $\arg\max_{W \in \mathcal{W}} G(W)$ is a singleton. Hence ϕ is single-valued. This completes the proof. □

Proof of Proposition 2(1). Let $C := \arg\max_{W \in \mathcal{W}} G(W)$. By Proposition 1, under Assumptions 1–2 we have $C = \phi(I_0)$, so C is non-empty. Fix any $I \in C$.

For any history h, let $A^-(h)$ be the set of agenda-setters whose turn has already occurred, including the current proposer if h is a voting history, and let $R(h) := I_0 \setminus A^-(h)$ be the set of players in I_0 who may still be selected as agenda-setters in future rounds.

Since $I \in \arg\max_{W \in \mathcal{W}} G(W)$, for every player $i \in I$ and every winning coalition $W \in \mathcal{W}$ with $i \in W$, Assumption 2 implies $U_i(I) \geq U_i(W)$, because $U_i(W) = x_i + g(i)G(W)$ and I maximizes G over $\mathcal{W}$. Moreover, since $I \in \mathcal{W} \setminus \{N\}$, Assumption 1(1) implies $U_i(I) = x_i + g(i)G(I) > x_i$ for every $i \in I$.

For every history h such that $R(h) \cap I = \emptyset$, the continuation game from h onward is a finite extensive-form game with complete information. Hence it admits a pure-strategy subgame-perfect equilibrium. Fix one such equilibrium and denote it by τ^h.

We now define a pure-strategy profile $\sigma^I = (\sigma_i^I)_{i \in N}$.

Strategy of player i.

Agenda-setting nodes. Suppose history h is an agenda-setting history and Nature selects player i as agenda-setter at h. Then

$$\sigma_i^I(h) = \begin{cases} \tau_i^h(h) & \text{if } R(h) \cap I = \emptyset, \\ I & \text{if } R(h) \cap I \neq \emptyset \text{ and } i \in I, \\ N & \text{if } R(h) \cap I \neq \emptyset \text{ and } i \notin I. \end{cases}$$

Voting nodes. Suppose history h is a voting history on a proposal $\mathcal{P} \subseteq N$, and player i is called upon to vote at h. Then

$$\sigma_i^I(h, \mathcal{P}) = \begin{cases} \tau_i^h(h, \mathcal{P}) & \text{if } R(h) \cap I = \emptyset, \\ \text{YES} & \text{if } R(h) \cap I \neq \emptyset,\ i \in I,\ \text{and } \mathcal{P} = I, \\ \text{NO} & \text{if } R(h) \cap I \neq \emptyset,\ i \in I,\ \text{and } \mathcal{P} \neq I, \\ \text{YES} & \text{if } R(h) \cap I \neq \emptyset \text{ and } i \notin I. \end{cases}$$

Step 1: subgames with no remaining agenda-setter from I. Consider any subgame beginning at a history h such that $R(h) \cap I = \emptyset$. By construction, the restriction of σ^I to the continuation game from h onward coincides with the fixed continuation equilibrium τ^h. Since τ^h is a subgame-perfect equilibrium of that continuation game, the restriction of σ^I to this subgame is a Nash equilibrium.

Step 2: subgames with at least one remaining agenda-setter from I. Now consider any subgame beginning at a history h such that $R(h) \cap I \neq \emptyset$.

We first show that no winning proposal $\mathcal{P} \neq I$ can be accepted in this subgame. Let $\mathcal{P}$ be any winning coalition proposed at some history h' in this subgame. Since both $\mathcal{P}$ and I are winning coalitions, $P_{\mathcal{P}} \geq \beta P_N$ and $P_I \geq \beta P_N$. Because $\beta > 1/2$, they cannot be disjoint: if $\mathcal{P} \cap I = \emptyset$, then $P_{\mathcal{P} \cup I} = P_{\mathcal{P}} + P_I \geq 2\beta P_N > P_N$, a contradiction. Therefore every winning proposal $\mathcal{P}$ satisfies $\mathcal{P} \cap I \neq \emptyset$.

Take any winning proposal $\mathcal{P} \neq I$ proposed at a history h' with $R(h') \cap I \neq \emptyset$. Since $\mathcal{P} \cap I \neq \emptyset$, some player $j \in \mathcal{P} \cap I$ must vote on $\mathcal{P}$, and because unanimity is required, $\mathcal{P}$ can be accepted only if j votes YES. Let h'' be the voting history at which player j is called upon to vote on $\mathcal{P}$. Since no additional agenda-setter is selected between h' and h'', we still have $R(h'') \cap I \neq \emptyset$. Hence, by construction, $\sigma_j^I(h'', \mathcal{P}) = \text{NO}$ whenever $\mathcal{P} \neq I$. Therefore no winning proposal different from I can be accepted while some member of I remains to be selected as an agenda-setter.

Next, since agenda-setters are drawn from I_0 without repetition and $R(h) \cap I \neq \emptyset$, if no proposal is accepted earlier then eventually some member of I is selected as agenda-setter. At that history she proposes I, and every member of I votes YES. Hence I is accepted and becomes the ruling coalition.

We now verify that no player has a profitable unilateral deviation in this subgame. Fix any player $i \in N$.

If $i \in I$, then under σ^I the eventual ruling coalition is I. Consider any unilateral deviation by i, either in her proposal or in her vote. Such a deviation may either leave the outcome equal to I, or lead to some eventual ruling coalition $W \in \mathcal{W}$. If $i \in W$, then Assumption 2 gives $U_i(W) = x_i + g(i)G(W)$. Since $I \in \arg\max_{W \in \mathcal{W}} G(W)$, we have $G(I) \geq G(W)$, and therefore $U_i(I) \geq U_i(W)$. If instead $i \notin W$, then Assumption 1(2) implies $U_i(W) < x_i$, whereas $U_i(I) > x_i$. Hence $U_i(I) > U_i(W)$. Thus no insider can profitably deviate.

If $i \notin I$, then any unilateral deviation by i leaves the outcome unchanged. Indeed, while $R(h) \cap I \neq \emptyset$, no winning proposal different from I can be accepted, and eventually some member of I is selected as agenda-setter, proposes I, and I is accepted. Hence every unilateral deviation by an outsider yields the same outcome and therefore the same payoff. So no outsider can profitably deviate.

Therefore the restriction of σ^I to any subgame with $R(h) \cap I \neq \emptyset$ is a Nash equilibrium.

Combining Steps 1 and 2, the restriction of σ^I to every subgame is a Nash equilibrium. Therefore σ^I is a subgame-perfect equilibrium. By construction, the ruling coalition induced by σ^I is I. This proves Proposition 2(1). □

Proof of Proposition 2(2). Assume $\beta \in (\frac{1}{2}, 1]$ and $\phi(I_0) = \{I\}$. Since $\phi(I_0) = \arg\max_{W \in \mathcal{W}} G(W)$ by Proposition 1, we have

$$G(I) > G(W) \quad \text{for all } W \in \mathcal{W} \setminus \{I\}. \tag{A.1}$$

For any $i \in I$ and any $W \in \mathcal{W}$ with $i \in W$, Assumption 2 gives $w_i(W) = g(i)G(W)$, hence

$$U_i(I) > U_i(W) \quad \text{for all } i \in I \text{ and all } W \in \mathcal{W} \setminus \{I\} \text{ with } i \in W. \tag{A.2}$$

Moreover, for any $W \in \mathcal{W} \setminus \{N\}$ and any $i \notin W$, Assumption 1(2) gives $U_i(W) = x_i + w_i(W) < x_i$. Finally, since $\beta > \frac{1}{2}$, any two winning coalitions intersect.

For any history h, let $R(h) \subseteq I_0$ be the set of remaining agenda-setters, and let $m(h) := |R(h)|$. We prove by backward induction on $m(h)$ that, in any subgame starting at h with $R(h) \cap I \neq \emptyset$, every SPE yields ruling coalition I.

If $m(h) = 1$, let a be the unique remaining agenda-setter. Since $R(h) \cap I \neq \emptyset$, we must have $a \in I$. Consider any SPE of the subgame at h.

First suppose that a proposes I. Let $i \in I$ be any player who is called upon to vote on I. If i deviates to NO, then I is rejected and the game moves to a continuation with no remaining agenda-setters, in which the eventual ruling coalition is some $J \in \mathcal{W}$. If

$i \in J$ and $J \neq I$, then $U_i(I) > U_i(J)$ by (A.2). If instead $i \notin J$, then Assumption 1(2) implies $U_i(J) < x_i$, whereas $U_i(I) > x_i$. Hence player i weakly prefers YES to NO. Since this holds for every member of I who is called upon to vote, and unanimity is required, the proposal I is accepted.

Now suppose instead that a proposes some $W \neq I$. If W is accepted and $a \in W$, then (A.2) implies $U_a(I) > U_a(W)$, so a can profitably deviate by proposing I instead. If W is accepted and $a \notin W$, then Assumption 1(2) implies $U_a(W) < x_a$, whereas $U_a(I) > x_a$, so again deviating to propose I is profitable. Finally, if W is rejected, then deviating to propose I yields acceptance of I and payoff $U_a(I)$. Therefore in any SPE, player a must propose I, and the ruling coalition is I.

Now fix $m \geq 2$ and assume the statement holds for all smaller values. Consider a subgame starting at h with $m(h) = m$ and $R(h) \cap I \neq \emptyset$, and fix any SPE. Suppose for contradiction that some $W \in \mathcal{W}$ with $W \neq I$ is accepted along the equilibrium path at some history $\tilde{h}$. Since both W and I are winning coalitions and $\beta > \frac{1}{2}$, they must overlap, so $W \cap I \neq \emptyset$. Let $i \in W \cap I$ be the first member of $W \cap I$ under the realized voting order who is called upon to vote on W at $\tilde{h}$.

Since unanimity is required, if i votes NO then W is rejected immediately. After this rejection, the current proposer is removed from the set of remaining agenda-setters, so the continuation begins at some history h' with $m(h') = m(\tilde{h}) - 1 < m$. We claim that $R(h') \cap I \neq \emptyset$. Indeed, if $R(h') \cap I = \emptyset$, then rejection must have removed the last member of I from the set of remaining agenda-setters. Since only the current proposer is removed when W is rejected, it would follow that the proposer at $\tilde{h}$ was the unique member of $R(\tilde{h}) \cap I$. But then $m(\tilde{h}) = 1$, contradicting $m \geq 2$. Hence $R(h') \cap I \neq \emptyset$.

Since the strategy profile under consideration is subgame perfect, its continuation after h' must constitute an SPE of the subgame starting at h'. By the induction hypothesis, that continuation yields ruling coalition I. It follows that, by deviating to NO, player i secures payoff $U_i(I)$, whereas by adhering to the prescribed action she receives $U_i(W)$. Because $i \in I \cap W$ and $W \neq I$, (A.2) implies $U_i(I) > U_i(W)$, contradicting sequential rationality at the voting node where i is called upon to vote on W. Hence no coalition $W \neq I$ can be accepted along the equilibrium path from h.

It follows that in any SPE from h, no coalition $W \neq I$ can be accepted along the equilibrium path. Hence the first accepted proposal, if any, must be I. When I is proposed, every member of I who is called upon to vote weakly prefers YES to NO by the same argument as in the base case, and unanimity therefore implies that I is accepted. Thus every SPE of the subgame at h yields ruling coalition I.

Finally, since $I_0 \in \mathcal{W}$ and $I \in \mathcal{W}$, and $\beta > \frac{1}{2}$, the two coalitions must overlap. Hence $I_0 \cap I \neq \emptyset$, so at the initial history h_0 we have $R(h_0) \cap I \neq \emptyset$. Applying the induction result at h_0, we conclude that every SPE of the full game yields ruling coalition I. $\square$

Proof of Proposition 3. We prove both directions.

($\Rightarrow$) Suppose $\phi(I_0) = \{I\}$. By Axiom 2, we have $I \in \mathcal{W}$. Since $\phi(I_0) = \{I\}$, Proposition 1 implies that I is the unique maximizer of G over $\mathcal{W}$. Hence $G(I) > G(W)$ for every $W \in \mathcal{W} \setminus \{I\}$.

Take any $A^{\text{ins}} \in (\mathcal{A}_I \setminus \{I\}) \cap \mathcal{W}$. Then $A^{\text{ins}} \in \mathcal{W}$ and $A^{\text{ins}} \neq I$, so the preceding inequality gives $G(I) > G(A^{\text{ins}})$. This proves condition (i).

Now take any $A^{\text{ext}} \in \mathcal{A}_{N\setminus I}$ and any $A^{\text{ins}} \in \mathcal{A}_I$ such that $A^{\text{ins}} \cup A^{\text{ext}} \in \mathcal{W}$. Since $A^{\text{ext}} \subseteq N \setminus I$ and $A^{\text{ext}} \neq \emptyset$, we have $A^{\text{ins}} \cup A^{\text{ext}} \neq I$. Therefore $A^{\text{ins}} \cup A^{\text{ext}} \in \mathcal{W} \setminus \{I\}$, and again the preceding inequality gives $G(I) > G(A^{\text{ins}} \cup A^{\text{ext}})$. This proves condition (ii).

($\Leftarrow$) Now suppose $I \in \mathcal{W}$ and conditions (i) and (ii) hold. We show that $\phi(I_0) = \{I\}$. By Proposition 1, it is enough to show that $G(I) > G(W)$ for all $W \in \mathcal{W} \setminus \{I\}$.

Fix any $W \in \mathcal{W} \setminus \{I\}$ and write $B := W \cap I$ and $C := W \setminus I$. Then $B \subseteq I$, $C \subseteq N \setminus I$, and $W = B \cup C$.

If $C = \emptyset$, then $W \subseteq I$. Since $W \in \mathcal{W}$ and $W \neq I$, there exists $\widetilde{A}^{\text{ins}} \in \mathcal{A}_I$ such that $P_{\widetilde{A}^{\text{ins}}} \geq P_W$ and $X_{\widetilde{A}^{\text{ins}}} \leq X_W$. Indeed, if $W \notin \mathcal{A}_I$, start from W and, whenever the current coalition is not in $\mathcal{A}_I$, replace it by a sub-coalition of I with strictly higher power and strictly lower resources. Since I is finite, this process terminates at some $\widetilde{A}^{\text{ins}} \in \mathcal{A}_I$. Because $P_{\widetilde{A}^{\text{ins}}} \geq P_W \geq \beta P_N$, we also have $\widetilde{A}^{\text{ins}} \in \mathcal{W}$, and since $W \neq I$ and $\widetilde{A}^{\text{ins}} \subseteq I$, necessarily $\widetilde{A}^{\text{ins}} \neq I$.

Condition (i) therefore gives $G(I) > G(\widetilde{A}^{\text{ins}})$. To compare $G(\widetilde{A}^{\text{ins}})$ and $G(W)$, note that $P_{\widetilde{A}^{\text{ins}}} \geq P_W$ and $X_{\widetilde{A}^{\text{ins}}} \leq X_W$. If $P_{\widetilde{A}^{\text{ins}}} = P_W$, then Assumption 1(1a) implies $G(\widetilde{A}^{\text{ins}}) \geq G(W)$. If $X_{\widetilde{A}^{\text{ins}}} = X_W$, then Assumption 1(1b) implies $G(\widetilde{A}^{\text{ins}}) \geq G(W)$. Otherwise, Assumption 1(1b) implies $G(P_{\widetilde{A}^{\text{ins}}}, X_W) \geq G(P_W, X_W) = G(W)$, and then Assumption 1(1a) implies $G(\widetilde{A}^{\text{ins}}) = G(P_{\widetilde{A}^{\text{ins}}}, X_{\widetilde{A}^{\text{ins}}}) \geq G(P_{\widetilde{A}^{\text{ins}}}, X_W)$. Hence $G(\widetilde{A}^{\text{ins}}) \geq G(W)$. Therefore $G(I) > G(W)$.

Now suppose $C \neq \emptyset$. As above, there exists $\widetilde{A}^{\text{ins}} \in \mathcal{A}_I$ such that $P_{\widetilde{A}^{\text{ins}}} \geq P_B$ and $X_{\widetilde{A}^{\text{ins}}} \leq X_B$. Similarly, since $N \setminus I$ is finite, there exists $\widetilde{A}^{\text{ext}} \in \mathcal{A}_{N\setminus I}$ such that $P_{\widetilde{A}^{\text{ext}}} \geq P_C$ and $X_{\widetilde{A}^{\text{ext}}} \leq X_C$.

Therefore $P_{\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}} = P_{\widetilde{A}^{\text{ins}}} + P_{\widetilde{A}^{\text{ext}}} \geq P_B + P_C = P_W$, and $X_{\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}} = X_{\widetilde{A}^{\text{ins}}} + X_{\widetilde{A}^{\text{ext}}} \leq X_B + X_C = X_W$. Since $W \in \mathcal{W}$, this implies $\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}} \in \mathcal{W}$. By condition (ii), $G(I) > G(\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}})$.

To compare $G(\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}})$ and $G(W)$, note that $P_{\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}} \geq P_W$ and $X_{\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}} \leq X_W$. If one of these inequalities is an equality, Assumption 1(1a) or Assumption 1(1b) gives $G(\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}) \geq G(W)$. Otherwise, Assumption 1(1b) implies $G(P_{\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}}, X_W) \geq G(P_W, X_W) = G(W)$, and then Assumption 1(1a) implies $G(\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}) \geq G(P_{\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}}, X_W)$.

Hence $G(\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}) \geq G(W)$. Therefore $G(I) > G(W)$.

Thus $G(I) > G(W)$ for every $W \in \mathcal{W} \setminus \{I\}$. Therefore I is the unique maximizer of G over $\mathcal{W}$, and Proposition 1 implies $\phi(I_0) = \{I\}$. □

Proof of Proposition 4. Let $I' := (I \setminus \{i,j\}) \cup \{i',j'\}$ be the post-exchange coalition, where $p_{i'} = p_i - \Delta p$, $x_{i'} = x_i + \Delta x$, $p_{j'} = p_j + \Delta p$, and $x_{j'} = x_j - \Delta x$, with $0 < \Delta p \leq (p_i - p_j)/2$ and $0 < \Delta x \leq (x_j - x_i)/2$. Then $p_{i'} \geq p_{j'}$ and $x_{i'} \leq x_{j'}$. Moreover, $P_{I'} = P_I$ and $X_{I'} = X_I$, so the indifference curve through the ruling coalition is unchanged. Denote this common indifference curve by $X = H_I(P)$.

For any sub-coalition $A \subseteq I$, define its shifted boundary by $H_A(P) := H_I(P + P_A) - X_A$. For any sub-coalition $A' \subseteq I'$, define $H_{A'}(P) := H_I(P + P_{A'}) - X_{A'}$. By Definition 5, the external safe areas of I and I' are $\mathcal{S}_I^{\text{ext}} = \bigcap_{A \in \mathcal{A}_I} \mathcal{S}_A^{\text{ext}}$ and $\mathcal{S}_{I'}^{\text{ext}} = \bigcap_{A' \in \mathcal{A}_{I'}} \mathcal{S}_{A'}^{\text{ext}}$, where $\mathcal{S}_A^{\text{ext}} := \{(P, X) \in \mathbb{R}^2_{++} : P < \beta P_N,\ X > H_A(P)\}$, and similarly for $\mathcal{S}_{A'}^{\text{ext}}$.

We begin with a comparison claim for best insider sub-coalitions.

Claim. For every $A' \in \mathcal{A}_{I'}$, there exists $A \in \mathcal{A}_I$ such that $P_A \geq P_{A'}$ and $X_A \leq X_{A'}$.

Fix $A' \in \mathcal{A}_{I'}$. We first construct a pre-exchange counterpart $A^- \subseteq I$ such that $P_{A^-} \geq P_{A'}$ and $X_{A^-} \leq X_{A'}$.

1. If $i', j' \in A'$, let $A^- := (A' \setminus \{i', j'\}) \cup \{i, j\}$. Then $P_{A^-} = P_{A'}$ and $X_{A^-} = X_{A'}$.
2. If $i', j' \notin A'$, let $A^- := A'$. Then again $P_{A^-} = P_{A'}$ and $X_{A^-} = X_{A'}$.
3. If $i' \in A'$ and $j' \notin A'$, let $A^- := (A' \setminus \{i'\}) \cup \{i\}$. Since $p_i = p_{i'} + \Delta p$ and $x_i = x_{i'} - \Delta x$, we have $P_{A^-} = P_{A'} + \Delta p > P_{A'}$ and $X_{A^-} = X_{A'} - \Delta x < X_{A'}$.
4. If $j' \in A'$ and $i' \notin A'$, let $A^- := (A' \setminus \{j'\}) \cup \{i\}$. Since $p_{j'} = p_j + \Delta p$ and $x_{j'} = x_j - \Delta x$, we have $P_{A^-} - P_{A'} = p_i - p_{j'} = p_i - p_j - \Delta p > 0$, because $0 < \Delta p \leq (p_i - p_j)/2$, and $X_{A^-} - X_{A'} = x_i - x_{j'} = x_i - x_j + \Delta x < 0$, because $0 < \Delta x \leq (x_j - x_i)/2$. Hence $P_{A^-} > P_{A'}$ and $X_{A^-} < X_{A'}$.

Thus, in every case, $P_{A^-} \geq P_{A'}$ and $X_{A^-} \leq X_{A'}$.

If $A^- \in \mathcal{A}_I$, set $A := A^-$. Otherwise, by Definition 3, there exists a coalition $B \subseteq I$ such that $P_B > P_{A^-}$ and $X_B < X_{A^-}$. If $B \in \mathcal{A}_I$, set $A := B$. If not, repeat the same argument with B in place of A^-. Since I is finite and each step strictly increases power while strictly decreases resources, this process terminates after finitely many steps at some $A \in \mathcal{A}_I$ satisfying $P_A \geq P_{A^-}$ and $X_A \leq X_{A^-}$. Combining this with $P_{A^-} \geq P_{A'}$ and $X_{A^-} \leq X_{A'}$ yields $P_A \geq P_{A'}$ and $X_A \leq X_{A'}$. This proves the claim.

We now show that the exchange weakly enlarges the external safe area.

Take any best sub-coalition of outsiders $B \in \mathcal{A}_{N\setminus I}$. Because the exchange only redistributes power and resources among members of the ruling coalition, it leaves the outsiders and their characteristics unchanged. Hence $N \setminus I = N \setminus I'$, and therefore $\mathcal{A}_{N\setminus I} = \mathcal{A}_{N\setminus I'}$.

Suppose $(P_B, X_B) \in \mathcal{S}_I^{\text{ext}}$. By Definition 5, this means that for every $A \in \mathcal{A}_I$ such that $A \cup B \in \mathcal{W}$, we have $X_B > H_A(P_B)$, or equivalently, $X_A + X_B > H_I(P_A + P_B)$.

We claim that $(P_B, X_B) \in \mathcal{S}_{I'}^{\text{ext}}$. Let $A' \in \mathcal{A}_{I'}$ be any insider best sub-coalition such that $A' \cup B \in \mathcal{W}$. By the claim just proved, there exists $A \in \mathcal{A}_I$ with $P_A \geq P_{A'}$ and $X_A \leq X_{A'}$. Since $A' \cup B \in \mathcal{W}$, we have $P_{A'} + P_B \geq \beta P_N$, hence $P_A + P_B \geq P_{A'} + P_B \geq \beta P_N$. Therefore $A \cup B \in \mathcal{W}$. Because $(P_B, X_B) \in \mathcal{S}_I^{\text{ext}}$, it follows that $X_A + X_B > H_I(P_A + P_B)$.

Since $P_A \geq P_{A'}$ and H_I is strictly increasing, we have $H_I(P_A + P_B) \geq H_I(P_{A'} + P_B)$. Moreover, $X_{A'} \geq X_A$. Therefore $X_{A'} + X_B \geq X_A + X_B > H_I(P_A + P_B) \geq H_I(P_{A'} + P_B)$. Equivalently, $X_B > H_{A'}(P_B)$. Since $A' \in \mathcal{A}_{I'}$ with $A' \cup B \in \mathcal{W}$ was arbitrary, we conclude that $(P_B, X_B) \in \mathcal{S}_{I'}^{\text{ext}}$. Hence $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{I'}^{\text{ext}}$.

It remains to verify that internal stability is preserved after the exchange. Let $A' \in \mathcal{A}_{I'} \cap \mathcal{W}$ be any winning best insider sub-coalition after the exchange. By the claim, there exists $A \in \mathcal{A}_I$ such that $P_A \geq P_{A'}$ and $X_A \leq X_{A'}$. Since $A' \in \mathcal{W}$, we have $P_{A'} \geq \beta P_N$, and therefore $P_A \geq \beta P_N$. Hence $A \in \mathcal{A}_I \cap \mathcal{W}$.

Because I was internally stable before the exchange, Proposition 3(i) implies $X_A > H_I(P_A)$. Moreover, since H_I is strictly increasing and $P_{A'} \leq P_A$, we have $H_I(P_{A'}) \leq H_I(P_A)$. Combining this with $X_{A'} \geq X_A$ yields $X_{A'} \geq X_A > H_I(P_A) \geq H_I(P_{A'})$, and therefore $X_{A'} > H_I(P_{A'})$. Thus A' lies in the internal safe area. Since $A' \in \mathcal{A}_{I'} \cap \mathcal{W}$ was arbitrary, no winning best insider sub-coalition can profitably secede after the exchange.

Therefore I' remains internally stable, and we have already shown that $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{I'}^{\text{ext}}$. Hence the exchange weakly increases the external resilience of the ruling coalition. □

Proof of Proposition 5. Fix a ruling coalition I, and let $H_I(\cdot)$ denote the indifference curve through (P_I, X_I), so that $X_I = H_I(P_I)$. Let $A^{\text{ins}} \in (\mathcal{A}_I \setminus \{I\}) \cap \mathcal{W}$ be any nontrivial winning best insider sub-coalition, and write $(P_A, X_A) := (P_{A^{\text{ins}}}, X_{A^{\text{ins}}})$.

Because $A^{\text{ins}} \subsetneq I$, we have $P_A < P_I$. Since I is a ruling coalition, Proposition 3(i) implies $G(I) > G(A^{\text{ins}})$. By the definition of H_I, this is equivalent to $X_A > H_I(P_A)$.

For any insider coalition $K \subseteq I$, define its shifted boundary by $H_I^K(P) := H_I(P + P_K) - X_K$, and let the corresponding external safe region be $\mathcal{S}_K^{\text{ext}} := \{(P, X) \in \mathbb{R}_{++}^2 : P < \beta P_N,\ X > H_I^K(P)\}$. In particular, $H_I^I(P) = H_I(P + P_I) - X_I$ and $H_I^A(P) = H_I(P + P_A) - X_A$.

We claim that $H_I^I(P) > H_I^A(P)$ for every $P \geq 0$ for which both expressions are well defined. Since $\mathcal{S}_K^{\text{ext}} = \{(P, X) \in \mathbb{R}_{++}^2 : \ P < \beta P_N,\ X > H_I^K(P)\}$ for each $K \subseteq I$, this inequality implies that every point in $\mathcal{S}_I^{\text{ext}}$ also belongs to $\mathcal{S}_{A^{\text{ins}}}^{\text{ext}}$. Hence $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{A^{\text{ins}}}^{\text{ext}}$.

To prove the claim, let $\Delta := P_I - P_A > 0$. Then $H_I^I(P) - H_I^A(P) = \big[H_I(P+P_A+\Delta) - H_I(P+P_A)\big] - (X_I - X_A)$. Let $t := P + P_A$. Since $P \geq 0$, we have $t \geq P_A$. Because H_I is strictly convex, for any fixed $\Delta > 0$ the increment $H_I(t+\Delta) - H_I(t)$ is strictly increasing in t. Therefore $H_I(P+P_A+\Delta) - H_I(P+P_A) \geq H_I(P_A+\Delta) - H_I(P_A) = H_I(P_I) - H_I(P_A)$.

Also, $X_I = H_I(P_I)$ and $X_A > H_I(P_A)$ imply $X_I - X_A < H_I(P_I) - H_I(P_A)$. Combining the two inequalities yields $H_I^I(P) - H_I^A(P) \geq \big[H_I(P_I) - H_I(P_A)\big] - (X_I - X_A) > 0$. This proves $H_I^I(P) > H_I^A(P)$ for every $P \geq 0$ in the common domain.

Since $A^{\text{ins}} \in (\mathcal{A}_I \setminus \{I\}) \cap \mathcal{W}$ was arbitrary, we obtain $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{A^{\text{ins}}}^{\text{ext}}$ for every nontrivial winning best insider sub-coalition. Because $I \in \mathcal{A}_I \cap \mathcal{W}$, it follows that $\bigcap_{A^{\text{ins}} \in \mathcal{A}_I \cap \mathcal{W}} \mathcal{S}_{A^{\text{ins}}}^{\text{ext}} = \mathcal{S}_I^{\text{ext}}$.

Finally, any internal reallocation that preserves (P_I, X_I) leaves the indifference curve through the ruling coalition unchanged, and hence leaves $H_I^I(P) = H_I(P + P_I) - X_I$ unchanged as well. Since the external safe area is exactly $\mathcal{S}_I^{\text{ext}}$, it follows that external resilience is unchanged. Therefore external resilience is invariant to any internal reallocation that preserves (P_I, X_I) and internal stability. □

Proof of Proposition 6. Fix a ruling coalition I and consider a finite sequence of within-I bilateral exchanges that preserves (P_I, X_I) and terminates at a hierarchical allocation $I' := I^T$. Let I^t denote the post-exchange coalition after step $t = 0, \ldots, T$, where $I^0 = I$. By Proposition 4, each exchange weakly enlarges the external safe area, so $\mathcal{S}_{I^0}^{\text{ext}} \subseteq \mathcal{S}_{I^1}^{\text{ext}} \subseteq \cdots \subseteq \mathcal{S}_{I^T}^{\text{ext}}$. Hence, if the inclusion is strict at any step, then $\mathcal{S}_I^{\text{ext}} \subsetneq \mathcal{S}_{I'}^{\text{ext}}$ follows immediately.

It therefore suffices to consider the case in which all earlier inclusions are equalities, that is, $\mathcal{S}_{I^t}^{\text{ext}} = \mathcal{S}_I^{\text{ext}}$ for every $t < T$, and to prove that the final exchange yields a strict enlargement: $\mathcal{S}_{I^{T-1}}^{\text{ext}} \subsetneq \mathcal{S}_{I^T}^{\text{ext}}$. We establish this strict inclusion first for the CES family G_ρ with $\rho < 1$ sufficiently small, and then pass to the Leontief limit. In particular, we show that there exists $\bar{\rho} < 1$ such that for every $\rho \in (\bar{\rho}, 1)$ sufficiently close to the Leontief limit, the final exchange shifts a relevant translated boundary strictly downward on a nonempty compact interval on which it uniquely determines the external envelope. This implies $\mathcal{S}_{I^{T-1}}^{\text{ext}}(G_\rho) \subsetneq \mathcal{S}_{I^T}^{\text{ext}}(G_\rho)$ for all such ρ, and continuity of the boundary family then yields $\mathcal{S}_{I^{T-1}}^{\text{ext}} \subsetneq \mathcal{S}_{I^T}^{\text{ext}}$ in the limit environment as well.

For the remainder of the proof, let $\hat{I} := I^{T-1}$ and $\hat{I}^1 := I^T$. Since $\hat{I}$ is the allocation immediately before the final exchange, it is not yet hierarchical. Hence there exist two members $i, j \in \hat{I}$ such that $p_i > p_j$ and $x_i < x_j$; that is, power and resources are not yet aligned across members of $\hat{I}$.

The final within-$\hat{I}$ exchange acts only on this pair. It transfers a small amount of power from i to j and a small amount of resources from j to i, so that $p_{i'} = p_i - \Delta p$, $x_{i'} = x_i + \Delta x$, $p_{j'} = p_j + \Delta p$, and $x_{j'} = x_j - \Delta x$, with $0 < \Delta p \leq (p_i - p_j)/2$ and

$0 < \Delta x \le (x_j - x_i)/2$. All other members of $\hat{I}$ are unchanged. The resulting allocation is denoted by $\hat{I}^1 := (\hat{I} \setminus \{i,j\}) \cup \{i', j'\}$. By construction, $P_{\hat{I}^1} = P_{\hat{I}}$ and $X_{\hat{I}^1} = X_{\hat{I}}$.

Step 1. Construction of a winning best sub-coalition A^* with $i \in A^*$ and $j \notin A^*$. Since $p_i/x_i > p_j/x_j$, Assumption 4 yields some $\lambda \in (p_j/x_j,\ p_i/x_i)$ such that $A^* := \{k \in \hat{I} : p_k - \lambda x_k > 0\}$ is winning. By the choice of λ, we have $p_i - \lambda x_i > 0$ and $p_j - \lambda x_j < 0$, hence $i \in A^*$ and $j \notin A^*$. In particular, $A^* \subsetneq \hat{I}$.

It remains to show that A^* is a best sub-coalition of $\hat{I}$. For any $A \subseteq \hat{I}$, define $\Phi_\lambda(A) := P_A - \lambda X_A = \sum_{k \in A}(p_k - \lambda x_k)$. Because Φ_λ is additively separable across members, a maximizer is obtained by including every player with positive contribution and excluding every player with negative contribution. Thus A^* maximizes Φ_λ over all subsets of $\hat{I}$.

Now suppose, toward a contradiction, that there exists some $B \subseteq \hat{I}$ such that $P_B > P_{A^*}$ and $X_B < X_{A^*}$. Then $\Phi_\lambda(B) - \Phi_\lambda(A^*) = (P_B - P_{A^*}) - \lambda(X_B - X_{A^*}) > 0$, because $P_B - P_{A^*} > 0$, $X_B - X_{A^*} < 0$, and $\lambda > 0$. This contradicts the maximality of A^*. Hence no subset of $\hat{I}$ has both strictly higher power and strictly lower resources than A^*, so $A^* \in \mathcal{A}_{\hat{I}}$.

Therefore A^* is a nontrivial winning best sub-coalition of $\hat{I}$, i.e. $A^* \in (\mathcal{A}_{\hat{I}} \setminus \{\hat{I}\}) \cap \mathcal{W}$.

Step 2. The final exchange and preservation of the winning status of the relevant bloc. Since A^* is strictly winning, $P_{A^*} - \beta P_N > 0$. Also, because $p_i > p_j$ and $x_i < x_j$, we have $(p_i - p_j)/2 > 0$ and $(x_j - x_i)/2 > 0$. Choose Δp and Δx such that $0 < \Delta p < \min\{(p_i - p_j)/2,\ P_{A^*} - \beta P_N\}$ and $0 < \Delta x < (x_j - x_i)/2$. Define the post-exchange characteristics by $p_{i'} = p_i - \Delta p$, $x_{i'} = x_i + \Delta x$, $p_{j'} = p_j + \Delta p$, and $x_{j'} = x_j - \Delta x$, and let $\hat{I}^1 := (\hat{I} \setminus \{i,j\}) \cup \{i', j'\}$. Then $p_{i'} > p_{j'}$ and $x_{i'} < x_{j'}$, so the pair remains ordered in the same direction after the exchange, but is strictly less misaligned than before.

Moreover, the exchange preserves the aggregate characteristics of the ruling coalition: $P_{\hat{I}^1} = P_{\hat{I}}$ and $X_{\hat{I}^1} = X_{\hat{I}}$. Hence, for every CES specification G_ρ under consideration, the indifference curve through the ruling coalition is the same before and after the exchange. We therefore denote this common curve by $H^\rho_{\hat{I}}$.

Let $A^\dagger \subseteq \hat{I}^1$ be the image of A^* under the exchange. Since $i \in A^*$ and $j \notin A^*$, the only change is that i is replaced by i'. Therefore $P_{A^\dagger} = P_{A^*} - \Delta p$ and $X_{A^\dagger} = X_{A^*} + \Delta x$. By the choice of Δp, $P_{A^\dagger} = P_{A^*} - \Delta p > \beta P_N$, so $A^\dagger$ remains winning. **Step 3. The translated boundary generated by A^* shifts strictly downward after the exchange.** For any sub-coalition $C \subseteq \hat{I}$ or $C \subseteq \hat{I}^1$, define its translated boundary by $H^\rho_C(P) := H^\rho_{\hat{I}}(P + P_C) - X_C$ for $P \in [0, \beta P_N)$, and let $\mathcal{S}^{\text{ext}}_C := \{(P, X) \in \mathbb{R}^2_{++} : P < \beta P_N,\ X > H^\rho_C(P)\}$.

Apply this definition to the pre-exchange bloc A^* and its post-exchange image $A^\dagger$. By Step 2, $P_{A^\dagger} = P_{A^*} - \Delta p < P_{A^*}$ and $X_{A^\dagger} = X_{A^*} + \Delta x > X_{A^*}$. Since $H^\rho_{\hat{I}}$ is strictly increasing, for every $P \in [0, \beta P_N)$ we have $H_{\hat{I}}^\rho(P + P_{A^\dagger}) < H_{\hat{I}}^\rho(P + P_{A^*})$. Subtracting the larger quantity $X_{A^\dagger}$ on the left and the smaller quantity X_{A^*} on the right yields

$$H_{A^\dagger}^\rho(P) = H_{\hat{I}}^\rho(P + P_{A^\dagger}) - X_{A^\dagger} < H_{\hat{I}}^\rho(P + P_{A^*}) - X_{A^*} = H_{A^*}^\rho(P)$$

for every $P \in [0, \beta P_N)$.

Thus the translated boundary associated with the relevant winning bloc shifts strictly downward at every outsider power level in the admissible domain. **Step 4. In the Leontief limit, A^* uniquely determines the pre-exchange finite envelope on a nonempty compact interval.**

Let $\mathcal{R} := (\mathcal{A}_{\hat{I}} \setminus \{\hat{I}\}) \cap \mathcal{W}$ be the set of nontrivial winning best sub-coalitions of $\hat{I}$. By Step 1, $A^* \in \mathcal{R}$. Since $A^* \subsetneq \hat{I}$, we have $0 < P_{\hat{I}} - P_{A^*}$. Since $A^* \in \mathcal{W}$, we also have $P_{A^*} \geq \beta P_N$, so $P_{\hat{I}} - P_{A^*} \leq P_N - \beta P_N = (1 - \beta) P_N < \beta P_N$. Thus the jump point associated with A^* lies strictly inside the outsider domain $(0, \beta P_N)$.

Leontief limit of translated boundaries. For the CES family $G_\rho(P, X) = \left[\alpha(P/P_{\hat{I}})^\rho + (1 - \alpha)((\bar{X} - X)/(\bar{X} - X_{\hat{I}}))^\rho\right]^{1/\rho}$, the indifference curve through $(P_{\hat{I}}, X_{\hat{I}})$ is defined implicitly by $\alpha(P/P_{\hat{I}})^\rho + (1 - \alpha)((\bar{X} - H_{\hat{I}}^\rho(P))/(\bar{X} - X_{\hat{I}}))^\rho = 1$. As $\rho \to -\infty$, the CES aggregator converges pointwise to the minimum. Hence the indifference curve through $(P_{\hat{I}}, X_{\hat{I}})$ converges pointwise to the inverse-L through that point: for $P < P_{\hat{I}}$ it is the horizontal line $X = X_{\hat{I}}$, while at $P = P_{\hat{I}}$ it has a vertical segment. Therefore, for each $C \in \mathcal{R}$, the translated boundary $H_C^\rho(P) := H_{\hat{I}}^\rho(P + P_C) - X_C$ converges pointwise on $[0, \beta P_N)$ to

$$H_C^{-\infty}(P) = \begin{cases} X_{\hat{I}} - X_C, & \text{if } P \leq P_{\hat{I}} - P_C, \\ +\infty, & \text{if } P > P_{\hat{I}} - P_C. \end{cases}$$

Staircase structure. Enumerate $\mathcal{R} = \{C_1, \ldots, C_m\}$ so that $P_{C_1} > \cdots > P_{C_m}$. Because each C_r is a best sub-coalition of $\hat{I}$, this ordering implies $X_{C_1} < \cdots < X_{C_m}$. Indeed, if $r < s$ and $X_{C_r} \geq X_{C_s}$, then C_r would have strictly higher power and weakly lower resources than C_s; by genericity, the weak inequality in resources must in fact be strict, so C_r would dominate C_s, contradicting $C_s \in \mathcal{A}_{\hat{I}}$.

Thus, as r increases, the jump points $P_{\hat{I}} - P_{C_r}$ strictly increase, while the finite horizontal levels $X_{\hat{I}} - X_{C_r}$ also strictly increase.

Let $A^* = C_q$. Since the jump points are strictly ordered, there exists $\varepsilon > 0$ such that $J := (P_{\hat{I}} - P_{A^*} - \varepsilon,\ P_{\hat{I}} - P_{A^*})$ contains no jump point of any coalition in $\mathcal{R} \setminus \{A^*\}$. On J, every coalition C_r with $r < q$ has already jumped, so $H_{C_r}^{-\infty}(P) = +\infty$ for all $P \in J$. Every coalition C_r with $r \geq q$ remains on its finite horizontal branch throughout J, and for every $r > q$ we have $X_{\hat{I}} - X_{C_r} < X_{\hat{I}} - X_{A^*} = H_{A^*}^{-\infty}(P)$. Hence among the translated boundaries that remain finite on J, the boundary of A^* is uniquely highest.

Define $\mathcal{R}^{\text{fin}} := \{C \in \mathcal{R} : P_{\hat{I}} - P_C \geq \sup J\}$, the set of coalitions whose limiting translated boundaries remain finite throughout J. Then $A^* \in \mathcal{R}^{\text{fin}}$, and for every $P \in J$ and every $C \in \mathcal{R}^{\text{fin}} \setminus \{A^*\}$, $H_{A^*}^{-\infty}(P) > H_C^{-\infty}(P)$.

Now fix any compact interval $K \subset J$. Since $\mathcal{R}^{\text{fin}}$ is finite and $H_{A^*}^{-\infty}$ strictly dominates every other finite limiting boundary on K, there exists $\eta > 0$ such that $H_{A^*}^{-\infty}(P) \geq H_C^{-\infty}(P) + \eta$ for all $P \in K$ and all $C \in \mathcal{R}^{\text{fin}} \setminus \{A^*\}$.

Uniform convergence on K for the finite branch family. For every $C \in \mathcal{R}^{\text{fin}}$ and every $P \in K$, we have $P + P_C < P_{\hat{I}}$. Since $\mathcal{R}^{\text{fin}}$ is finite and K is compact, there exists $\delta > 0$ such that $P + P_C \leq P_{\hat{I}} - \delta$ for every $P \in K$ and every $C \in \mathcal{R}^{\text{fin}}$. Thus all evaluation points stay a fixed positive distance to the left of the kink.

For the CES family, the indifference curve $H_{\hat{I}}^{\rho}$ through $(P_{\hat{I}}, X_{\hat{I}})$ converges, as $\rho \to -\infty$, to the Leontief inverse-L through $(P_{\hat{I}}, X_{\hat{I}})$. Since $K + \{P_C : C \in \mathcal{R}^{\text{fin}}\}$ is contained in the compact interval $[0, P_{\hat{I}} - \delta]$, this convergence is uniform on that set. Hence $H_{\hat{I}}^{\rho}(P + P_C) \to X_{\hat{I}}$ uniformly over $P \in K$ and $C \in \mathcal{R}^{\text{fin}}$.

Therefore $H_C^{\rho}(P) = H_{\hat{I}}^{\rho}(P + P_C) - X_C \to X_{\hat{I}} - X_C = H_C^{-\infty}(P)$ uniformly on K, uniformly over $C \in \mathcal{R}^{\text{fin}}$. Since $\mathcal{R}^{\text{fin}}$ is finite, this convergence is joint.

It follows that there exists $\bar{\rho} < 0$ such that, for every $\rho \leq \bar{\rho}$, every $P \in K$, and every $C \in \mathcal{R}^{\text{fin}}$, we have $|H_C^{\rho}(P) - H_C^{-\infty}(P)| < \eta/3$. Hence, for every $\rho \leq \bar{\rho}$, every $P \in K$, and every $C \in \mathcal{R}^{\text{fin}} \setminus \{A^*\}$, $H_{A^*}^{\rho}(P) \geq H_C^{\rho}(P) + \eta/3$. Thus, for all sufficiently negative ρ, the finite pre-exchange envelope on K is uniquely determined by $H_{A^*}^{\rho}$.

Step 5. The post-exchange envelope lies strictly below the pre-exchange envelope on K. Let $\mathcal{R}_{\hat{I}^1} := (\mathcal{A}_{\hat{I}^1} \setminus \{\hat{I}^1\}) \cap \mathcal{W}$ denote the set of nontrivial winning best sub-coalitions after the exchange. By Proposition 4, for every $B' \in \mathcal{R}_{\hat{I}^1}$ there exists an antecedent coalition $B \in \mathcal{R}$ such that $P_B \geq P_{B'}$ and $X_B \leq X_{B'}$. Since $H_{\hat{I}}^{\rho}$ is strictly increasing, this implies $H_{B'}^{\rho}(P) \leq H_B^{\rho}(P)$ for every $P \in [0, \beta P_N)$.

Fix $P \in K$ and let $B' \in \mathcal{R}_{\hat{I}^1}$ be arbitrary, with antecedent $B \in \mathcal{R}$.

If $B \neq A^*$, then Step 4 implies $H_B^{\rho}(P) \leq H_{A^*}^{\rho}(P) - \eta/3$. Hence $H_{B'}^{\rho}(P) \leq H_B^{\rho}(P) < H_{A^*}^{\rho}(P)$.

If $B = A^*$, then B' is a post-exchange image of A^*. By the antecedent comparison, we have $P_{B'} \leq P_{A^*}$ and $X_{B'} \geq X_{A^*}$. Moreover, by the generic choice of $(\Delta p, \Delta x)$ in the admissible open set, these inequalities cannot both hold with equality. Therefore at least one is strict. Since $H_{\hat{I}}^{\rho}$ is strictly increasing, it follows that $H_{\hat{I}}^{\rho}(P + P_{B'}) \leq H_{\hat{I}}^{\rho}(P + P_{A^*})$, with strict inequality if $P_{B'} < P_{A^*}$, while $-X_{B'} \leq -X_{A^*}$, with strict inequality if $X_{B'} > X_{A^*}$. Thus in this case as well, $H_{B'}^{\rho}(P) < H_{A^*}^{\rho}(P)$.

Therefore, for every $P \in K$ and every $B' \in \mathcal{R}_{\hat{I}1}$, we have $H^\rho_{B'}(P) < H^\rho_{A^*}(P)$. Since Step 4 shows that $H^\rho_{A^*}$ uniquely determines the pre-exchange envelope on K, it follows that $\max_{B' \in \mathcal{R}_{\hat{I}1}} H^\rho_{B'}(P) < H^\rho_{A^*}(P) = \max_{B \in \mathcal{R}} H^\rho_B(P)$ for every $P \in K$.

Step 6. Strict expansion of the external safe area. Fix any $P \in K$. By Step 5, $\max_{B' \in \mathcal{R}_{\hat{I}1}} H^\rho_{B'}(P) < \max_{B \in \mathcal{R}} H^\rho_B(P)$. Choose any X strictly between these two values. Then $X > \max_{B' \in \mathcal{R}_{\hat{I}1}} H^\rho_{B'}(P)$, so $(P, X) \in \mathcal{S}^{\text{ext}}_{\hat{I}1}$, while $X < \max_{B \in \mathcal{R}} H^\rho_B(P)$, so $(P, X) \notin \mathcal{S}^{\text{ext}}_{\hat{I}}$. Hence $(P, X) \in \mathcal{S}^{\text{ext}}_{\hat{I}1} \setminus \mathcal{S}^{\text{ext}}_{\hat{I}}$, which proves $\mathcal{S}^{\text{ext}}_{\hat{I}} \subsetneq \mathcal{S}^{\text{ext}}_{\hat{I}1}$ for every $\rho \leq \bar{\rho}$.

Since all earlier exchanges weakly enlarge the external safe area, we conclude that $\mathcal{S}^{\text{ext}}_{I} \subsetneq \mathcal{S}^{\text{ext}}_{I'}$ for every $\rho \leq \bar{\rho}$. □

Proof of Proposition 7. Fix a ruling coalition I, and let H and $\widetilde{H}$ be two indifference curves through (P_I, X_I) such that $H \succ \widetilde{H}$, that is, $H'(P) < \widetilde{H}'(P)$ for all relevant P.

Since both curves pass through (P_I, X_I), we have $H(P_I) = \widetilde{H}(P_I) = X_I$. Hence, for any $P < P_I$, $H(P) - \widetilde{H}(P) = \int_P^{P_I} (\widetilde{H}'(t) - H'(t))\, dt > 0$, so $H(P) > \widetilde{H}(P)$. Similarly, for any $P > P_I$ at which both curves are defined, $\widetilde{H}(P) - H(P) = \int_{P_I}^{P} (\widetilde{H}'(t) - H'(t))\, dt > 0$, so $H(P) < \widetilde{H}(P)$.

We first compare internal resilience. Any insider sub-coalition $A \subseteq I$ satisfies $P_A \leq P_I$, so only the region $P \leq P_I$ is relevant. Under H, the internal safe area is $\mathcal{S}^{\text{int}}_I(H) := \{(P, X) : X > H(P)\}$, while under $\widetilde{H}$ it is $\mathcal{S}^{\text{int}}_I(\widetilde{H}) := \{(P, X) : X > \widetilde{H}(P)\}$. Since $H(P) > \widetilde{H}(P)$ for every $P < P_I$, we obtain $\mathcal{S}^{\text{int}}_I(H) \subseteq \mathcal{S}^{\text{int}}_I(\widetilde{H})$. The inclusion is strict. Fix any $P_0 < P_I$, and choose X_0 such that $\widetilde{H}(P_0) < X_0 < H(P_0)$. Then $(P_0, X_0) \in \mathcal{S}^{\text{int}}_I(\widetilde{H})$ but $(P_0, X_0) \notin \mathcal{S}^{\text{int}}_I(H)$. Hence $\mathcal{S}^{\text{int}}_I(H) \subsetneq \mathcal{S}^{\text{int}}_I(\widetilde{H})$, so internal resilience under H is strictly lower than under $\widetilde{H}$.

We next compare external resilience. Since both indifference curves are strictly convex, Proposition 5 applies. Therefore the external safe area is determined by the full coalition I: $\mathcal{S}^{\text{ext}}_I(H) = \{(P, X) : P < \beta P_N,\ X > H(P + P_I) - X_I\}$ and $\mathcal{S}^{\text{ext}}_I(\widetilde{H}) = \{(P, X) : P < \beta P_N,\ X > \widetilde{H}(P + P_I) - X_I\}$. For every relevant $P > 0$, we have $P + P_I > P_I$. Hence $H(P + P_I) < \widetilde{H}(P + P_I)$. Subtracting X_I from both sides gives $H(P + P_I) - X_I < \widetilde{H}(P + P_I) - X_I$, and therefore $\mathcal{S}^{\text{ext}}_I(\widetilde{H}) \subseteq \mathcal{S}^{\text{ext}}_I(H)$.

This inclusion is also strict. Fix any $P_1 \in (0, \beta P_N)$, and choose X_1 such that $H(P_1 + P_I) - X_I < X_1 < \widetilde{H}(P_1 + P_I) - X_I$. Then $(P_1, X_1) \in \mathcal{S}^{\text{ext}}_I(H)$ but $(P_1, X_1) \notin \mathcal{S}^{\text{ext}}_I(\widetilde{H})$. Thus $\mathcal{S}^{\text{ext}}_I(\widetilde{H}) \subsetneq \mathcal{S}^{\text{ext}}_I(H)$, so external resilience under H is strictly higher than under $\widetilde{H}$.

Therefore, if $H \succ \widetilde{H}$, then the ruling coalition I has strictly lower internal resilience and strictly higher external resilience under H than under $\widetilde{H}$. □

Appendix B Examples

The following example illustrates a typical function $G_i(\cdot)$ satisfying Assumptions 1–3 and clarifies how the plundering technology shapes the relative value of power and internal resources.

Example 1. *For any* $i \in I \in \mathcal{W} \setminus \{N\}$*, consider*

$$w_i(I) = G_i(I) := \left(\frac{p_i}{P_I}\right)\left(\frac{P_I}{P_N}\right)^{\alpha+1}\left(\frac{X_N}{X_I}\right), \tag{B.1}$$

where $\alpha > 0$*. The term* $\frac{p_i}{P_I}$ *is player* i*'s share of plundered resources, proportional to her relative power in the ruling coalition, while the plunder function* $\left(\frac{P_I}{P_N}\right)^{\alpha+1}\left(\frac{X_N}{X_I}\right)$ *captures how aggregate coalition power and internal resources affect total extraction. One can verify that* $\alpha > 0$ *is required for parts (i)–(ii) of Assumption 1; if* $\alpha < 0$*, these conditions fail.*

Normalize $P_N = X_N = 1$*. Then*

$$G_i(I) = p_i P_I^{\alpha} \frac{1}{X_I}.$$

Fix a payoff level $\overline{G}_i$ *and let* I *be a ruling coalition containing* i*. The indifference curve of player* i *through* I *is the locus of* (P, X) *such that* $G_i(P, X) = \overline{G}_i$*:*

$$X = C_i(I) P^{\alpha}, \tag{B.2}$$

where $C_i(I) := \frac{p_i}{\overline{G}_i}$ *and* $P \in [\beta, 1]$*. Along such an indifference curve, the marginal rate of substitution between power and resources is*

$$MRS_{PX} = \alpha \frac{X}{P}.$$

The parameter α *governs the relative valuation of power versus internal resources. A higher* α *makes payoffs more sensitive to coalition power: for a given level of internal resources, an increase in* P_I *raises* $G_i(I)$ *more strongly. Equivalently, along an indifference curve, the marginal value of power relative to internal resources is higher when* α *is larger. In this sense, a higher* α *corresponds to a more power-intensive plundering technology.*

Conversely, lower values of α *reduce the marginal contribution of additional power to plunder. This corresponds to a weaker power effect in plundering: coalition payoffs become less responsive to power and relatively more constrained by internal resources. In that case, the trade-off between bringing in additional power and protecting more insiders'*

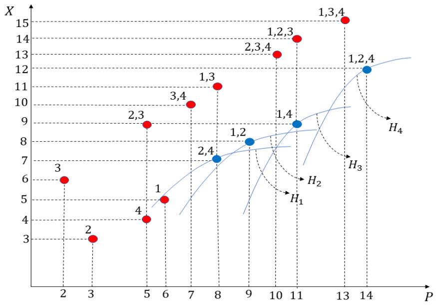

Figure 14: Ruling coalition under different indifference curves.

resources is less tilted toward power.

The next example shows that, without further restrictions on the joint distribution of power and resources and on the plundering function, there is no general characterization of the ruling coalition's composition. This follows, first, from Proposition 2(1), which implies that any coalition in the set of potential ruling coalitions may be the ruling coalition for some range of indifference curves; and second, from the fact that the set of potential ruling coalitions itself cannot be sharply characterized without additional structure on $(p_i, x_i)_{i \in N}$. In particular, there is no guarantee that the ruling coalition contains the most powerful player, the player with the fewest resources, or the player with the highest power-to-resource ratio (Figure 14).[10] Example 5 illustrates this point.

Example 2. *Suppose* $N = \{1, 2, 3, 4\}$, *with*

$$p_1 = 6, \; x_1 = 5, \quad p_2 = 3, \; x_2 = 3, \quad p_3 = 2, \; x_3 = 6, \quad p_4 = 5, \; x_4 = 4,$$

and let $\beta = \frac{1}{2} + \epsilon$. *Then the set of winning coalitions is*

$$\mathcal{W} = \{\{1, 2\}, \{1, 4\}, \{2, 4\}, \{1, 2, 4\}\}.$$

As shown in Figure 14, different indifference curves select different ruling coalitions: when the indifference curve is H_1, *the ruling coalition is* $\{2, 4\}$, *which excludes the player with the highest power (player 1); when it is* H_3, *the ruling coalition is* $\{1, 4\}$, *which excludes the player with the lowest resources; and when it is* H_2, *the ruling coalition excludes the player with the highest power-to-resource ratio (player 4). Thus, absent additional structure, the ruling coalition need not contain the most powerful player, the poorest player, or the player with the highest power-to-resource ratio.*

[10]Example 4 in the appendix shows that a sharper characterization is possible when powers, resources, or both are equally distributed in society.

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2008): "Coalition formation in non-democracies," *The Review of Economic Studies*, 75 (4), 987–1009.

——— (2012): "Dynamics and stability of constitutions, coalitions, and clubs," *American Economic Review*, 102 (4), 1446–1476.

Acemoglu, Daron and James A Robinson (2008): "Persistence of power, elites, and institutions," .

——— (2013): "Economics versus politics: Pitfalls of policy advice," *Journal of Economic perspectives*, 27 (2), 173–192.

Alatas, Syed Farid (2014): *Applying Ibn Khaldūn: The recovery of a lost tradition in sociology*, Routledge.

Andolfatto, David (2002): "A theory of inalienable property rights," *Journal of Political Economy*, 110 (2), 382–393.

Bai, Ying and Titi Zhou (2019): ""Mao's last revolution"': a dictator's loyalty-competence tradeoff," *Public Choice*, 180 (3), 469–500.

Battaglini, Marco (2021): "Coalition formation in legislative bargaining," *Journal of Political Economy*, 129 (11), 3206–3258.

Besley, Timothy and Torsten Persson (2009): "The origins of state capacity: Property rights, taxation, and politics," *American economic review*, 99 (4), 1218–1244.

Besley, Timothy J and Masayuki Kudamatsu (2007): "Making autocracy work," *LSE STICERD Research Paper No. DEDPS48*.

Brewer, Tony (2002): *Marxist theories of imperialism: A critical survey*, Routledge.

Dal Bó, Ernesto and Pedro Dal Bó (2011): "Workers, warriors, and criminals: social conflict in general equilibrium," *Journal of the European Economic Association*, 9 (4), 646–677.

Dal Bó, Ernesto, Pablo Hernández-Lagos, and Sebastián Mazzuca (2022): "The paradox of civilization: Preinstitutional sources of security and prosperity," *American Political Science Review*, 116 (1), 213–230.

De Mesquita, Ethan Bueno (2010): "Regime change and revolutionary entrepreneurs," *American Political Science Review*, 104 (3), 446–466.

Diermeier, Daniel, Georgy Egorov, and Konstantin Sonin (2017): "Political economy of redistribution," *Econometrica*, 85 (3), 851–870.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin (2009): "Why resource-poor dictators allow freer media: A theory and evidence from panel data," *American Political Science Review*, 103 (4), 645–668.

Egorov, Georgy and Konstantin Sonin (2011): "Dictators and their viziers: Endogenizing the loyalty–competence trade-off," *Journal of the European Economic Association*, 9 (5), 903–930.

——— (2024): "The political economics of non-democracy," *Journal of Economic Literature*, 62 (2), 594–636.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova (2020): "Social media and protest participation: Evidence from Russia," *Econometrica*, 88 (4), 1479–1514.

Garicano, Luis (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of political economy*, 108 (5), 874–904.

Garicano, Luis and Esteban Rossi-Hansberg (2015): "Knowledge-based hierarchies: Using organizations to understand the economy," *economics*, 7 (1), 1–30.

Gehlbach, Scott and Konstantin Sonin (2014): "Government control of the media," *Journal of public Economics*, 118, 163–171.

Guriev, Sergei and Daniel Treisman (2019): "Informational autocrats," *Journal of economic perspectives*, 33 (4), 100–127.

Hafer, Catherine (2006): "On the origins of property rights: Conflict and productionin the state of nature," *The Review of Economic Studies*, 73 (1), 119–143.

Harsanyi, John C. (1974): "An equilibrium-point interpretation of stable sets and a proposed alternative definition," *Management Science*, 20 (11), 1472–1495.

Henn, Soeren J, Christian Mastaki Mugaruka, Miguel Ortiz, Raúl Sánchez de la Sierra, and David Qihang Wu (2024): "Monopoly of Taxation without a Monopoly of Violence: The Weak State's Trade-Offs from Taxation," *Review of Economic Studies*, rdae036.

Jia, Ruixue, Masayuki Kudamatsu, and David Seim (2015): "Political selection in China: The complementary roles of connections and performance," *Journal of the European Economic Association*, 13 (4), 631–668.

KHALDUN, IBN (1967): "The Muqaddimah, an Introduction to History, tr. by Franz Rosenthal, 3 vols," .

KÖNIG, MICHAEL D, DOMINIC ROHNER, MATHIAS THOENIG, AND FABRIZIO ZILIBOTTI (2017): "Networks in conflict: Theory and evidence from the great war of africa," *Econometrica*, 85 (4), 1093–1132.

KONRAD, KAI A AND STERGIOS SKAPERDAS (2007): "Succession rules and leadership rents," *Journal of Conflict Resolution*, 51 (4), 622–645.

KURAN, TIMUR (1996): "The discontents of Islamic economic morality," *The American Economic Review*, 86 (2), 438–442.

LORENTZEN, PETER (2014): "China's strategic censorship," *American Journal of political science*, 58 (2), 402–414.

LORENTZEN, PETER L (2013): "Regularizing rioting: Permitting public protest in an authoritarian regime," *Quarterly Journal of Political Science*, 8 (2), 127–158.

LOWES, SARA, NATHAN NUNN, JAMES A ROBINSON, AND JONATHAN L WEIGEL (2017): "The evolution of culture and institutions: Evidence from the Kuba Kingdom," *Econometrica*, 85 (4), 1065–1091.

MATTINGLY, DANIEL C (2024): "How the party commands the gun: The foreign–domestic threat dilemma in China," *American Journal of Political Science*, 68 (1), 227–242.

MENG, ANNE (2020): *Constraining dictatorship*, Cambridge University Press Cambridge.

MICHALOPOULOS, STELIOS AND ELIAS PAPAIOANNOU (2013): "Pre-colonial ethnic institutions and contemporary African development," *Econometrica*, 81 (1), 113–152.

MOOKHERJEE, DILIP (2013): "Incentives in hierarchies," *The Handbook of Organizational Economics*, 764–798, edited by Robert Gibbons and John Roberts.

MORISHIMA, MICHIO (1974): "Marx in the light of modern economic theory," *Econometrica: Journal of the Econometric Society*, 611–632.

PAINE, JACK (2021): "The Dictator's Power-Sharing Dilemma: Countering Dual Outsider Threats," *American Journal of Political Science*, 65 (2), 510–527.

PERSSON, TORSTEN (2002): "Do political institutions shape economic policy?" *Econometrica*, 70 (3), 883–905.

Puga, Diego and Daniel Trefler (2014): "International trade and institutional change: Medieval Venice's response to globalization," *The Quarterly Journal of Economics*, 129 (2), 753–821.

Qian, Yingyi (1994): "Incentives and loss of control in an optimal hierarchy," *The Review of Economic Studies*, 61 (3), 527–544.

Qian, Yingyi, Gerard Roland, and Chenggang Xu (2006): "Coordination and experimentation in M-form and U-form organizations," *Journal of Political Economy*, 114 (2), 366–402.

Ray, Debraj and Rajiv Vohra (2015a): "Coalition Formation," in *Handbook of Game Theory with Economic Applications*, ed. by H. Peyton Young and Shmuel Zamir, North Holland, 240–322.

——— (2015b): "Coalition formation," *Handbook of game theory with economic applications*, 4, 239–326.

——— (2015c): "The farsighted stable set," *Econometrica*, 83 (3), 977–1011.

Roemer, John E (1980): "A general equilibrium approach to Marxian economics," *Econometrica: Journal of the Econometric Society*, 505–530.

——— (1981): *Analytical foundations of Marxian economic theory*, Cambridge University Press.

Sánchez De La Sierra, Raúl (2020): "On the origins of the state: Stationary bandits and taxation in eastern congo," *Journal of Political Economy*, 128 (1), 000–000.

Shadmehr, Mehdi (2018): "Multiplicity and Uniqueness in Generalized Regime Change Games," *Available at SSRN 3121241*.

Svolik, Milan W (2012): *The politics of authoritarian rule*, Cambridge University Press.

Wintrobe, Ronald (1990): "The tinpot and the totalitarian: An economic theory of dictatorship," *American political science review*, 84 (3), 849–872.

——— (2000): *The political economy of dictatorship*, Cambridge University Press.

Xu, Guo (2018): "The costs of patronage: Evidence from the british empire," *American Economic Review*, 108 (11), 3170–3198.

Yanagizawa-Drott, David (2014): "Propaganda and conflict: Evidence from the Rwandan genocide," *The Quarterly Journal of Economics*, 129 (4), 1947–1994.

Zakharov, Alexei V (2016): "The loyalty-competence trade-off in dictatorships and outside options for subordinates," *The Journal of Politics*, 78 (2), 457–466.