Plundering Coalitions*

Weijia Li† Alireza Mohammadi Sepahvand‡

November 26, 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 to the power and resources of both insiders and outsiders, as well as to the intensity of plundering. We show that hierarchy—an internal ordering in which members are equal within ranks and distinct ranks are weakly ordered in both power and resources, with at least one strict inequality—weakly increases external resilience, that is, resilience to shocks to outsiders' power and resources. The gain is strict when a relevant insider sub-coalition is locally decisive; a sufficiently low uniform bound on the marginal return to power guarantees local decisiveness. By contrast, when the marginal return to power is uniformly high, external resilience depends only on the ruling coalition's aggregate power and resources and is therefore unaffected by redistribution among its members. We interpret the low and the high marginal return to power as capturing weak and strong property-rights protection, respectively. Our final result identifies a condition for a trade-off between internal and external resilience.

Keywords: Political Economy, Coalition Formation, Institutions, Resilience, Plundering.
JEL Classifications: D72, D74, C72.

*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. ChatGPT was used to check mathematical arguments, develop and verify the numerical illustration, and prepare targeted revisions to wording and LaTeX source. The authors remain responsible for all content.

†Monash University, Australia. E-mail: Weijia.li@monash.edu.
‡University of Edinburgh, UK. E-mail: amohamma@ed.ac.uk.

1 Introduction

For most of recorded history, societies have been ruled by coalitions that live at the expense of outsiders (Olson, 1993; North et al., 2009). Military aristocracies lived on the tribute of peasants, privileged orders were exempt from the taxes that others paid, and armed groups still tax civilians in parts of the world today (İnalcık, 1973; Tocqueville, 1856; Sánchez de la Sierra, 2020; Henn et al., 2025). A coalition of this kind faces a trade-off that coalitions formed for production do not: each new member adds to its power, but the new member's own resources can no longer be taken. We call such a coalition a *plundering coalition* and ask how the distribution of power and resources among its members determines whether it survives. Our answer is that alignment protects: a plundering coalition is more resilient to outside challenges when those of its members who hold more power also hold more resources.

The internal distribution matters because outsiders cannot rule alone. A ruling coalition holds a supermajority of society's power, so any challenging coalition must include some of its members. Consider a coalition that includes a general, who commands many soldiers but owns little, and a landowner, who owns much but commands few. To a group of outsiders, the general is the ideal partner, bringing the force needed to win and little wealth that a new coalition would have to leave untouched. The landowner is the ideal victim, whose exclusion costs little force and exposes much wealth. Now let the general give the landowner some soldiers in exchange for some land, leaving the coalition's total force and wealth unchanged and the two members' order intact. The general becomes a costlier partner and the landowner a less tempting victim. The landowner becomes a somewhat better partner, but never a better one than the general was before the exchange. A challenge that was profitable before the exchange may therefore no longer be, and no group of outsiders acquires a profitable challenge that it did not have before.

Our main result shows that this logic is general. Call two members *mis-ordered* if one has more power and fewer resources than the other, and call an allocation of power and resources *hierarchical* if no two members are mis-ordered. Equivalently, the members of a hierarchical coalition can be divided into ranks: members of the same rank hold the same power and resources, and a higher rank holds weakly more power and weakly more resources than a lower rank, and strictly more of at least one. We show that any allocation that is not hierarchical can be turned, by a finite sequence of exchanges like the one above, into a hierarchical allocation with the same aggregate power and resources that survives every shock to the distribution of power and resources among outsiders that the original allocation survives. The comparison holds for every plundering technology in our class and for every probability distribution of shocks, and no exchange in the sequence gives any group of insiders an incentive to secede.[1] Hierarchy thus strengthens a coalition against outsiders at no cost to its internal cohesion.

We develop the argument in a model of coalition formation. A society consists of finitely many individuals, each endowed with power and resources. A coalition is winning if its aggregate power reaches a supermajority of society's power.[2] The ruling coalition plunders outsiders and divides the spoils among its members, whose own resources are protected. A member's gain rises with the coalition's aggregate power, which determines how much it can extract, and

[1] Each exchange must preserve resource feasibility, so that no coalition plunders more than the resources outside it. The requirement holds, for example, whenever each coalition's total plunder is a share of the resources outside it. We discuss this further in Section 4.

[2] Formally, a coalition is winning if its aggregate power is at least a fraction $\beta \in (\frac{1}{2}, 1)$ of society's total power. The supermajority requirement refers to power, not to votes.

falls with the coalition's aggregate resources, which cannot be plundered. Coalitions form through an adaptation of the protocol of Acemoglu et al. (2008): members of an initial winning coalition propose alternative coalitions in turn, and a proposal succeeds if the proposed coalition is winning and all of its members accept it. The equilibrium ruling coalition is the winning coalition that best balances power against protected resources, and it is unique under a no-ties condition.

A ruling coalition must withstand two kinds of threat: *secession*, by a winning subset of its own members, and *external deviations*, by coalitions that combine some of its members with outsiders (Proposition 3). Because outsiders cannot rule alone, every external threat recruits insiders. We call resilience to the first kind of threat *internal resilience* and resilience to the second kind *external resilience*. We measure each by a *safe area*: the set of aggregate power and resources at which a potential challenger poses no threat. The external safe area has a direct interpretation. Compare two allocations of the same aggregate power and resources among the same members, neither of which invites secession. If the external safe area under one contains that under the other, the coalition survives under the first allocation every shock to the distribution of power and resources among outsiders that it survives under the second. It is therefore at least as likely to survive under the first allocation, whatever the probability distribution of shocks, even when the endowments of different outsiders are arbitrarily correlated.

Our main result formalizes the logic of the example. An *exchange* between two mis-ordered members moves power from the member with more power to the other, resources in the opposite direction, or both, without reversing either ordering, and so leaves the coalition's aggregate power and resources unchanged. Every coalition that includes the more powerful member but not the other loses power or gains resources and becomes less threatening. Every coalition that includes the less powerful member but not the other is no more threatening than the same coalition with the more powerful member before the exchange. Hence every external deviation possible after the exchange is weakly less threatening than one possible before it, and the external safe area weakly expands (Proposition 4). A finite sequence of exchanges removes all mis-ordered pairs and therefore reaches a hierarchical allocation whose external safe area contains that of the original allocation (Corollary 1).

The hierarchical allocation that a sequence of exchanges reaches is not necessarily an equal allocation: the sequence stops as soon as no two members are mis-ordered, and the ranks that remain may be far apart. The equal allocation is only the special case with a single rank, and our result does not single it out, because two hierarchical allocations with the same aggregate power and resources cannot in general be compared in terms of external resilience. Finally, the result does not require members to rank coalitions in the same way. When members rank coalitions differently, each exchange still weakly expands the external safe area, and so does the whole sequence, provided that the two members involved in each exchange do not differ too much in their willingness to join alternative coalitions (Proposition 6).

How large is the gain from hierarchy? It depends on the *marginal return to power*: the additional protected resources that one more unit of power compensates for. In choosing partners among the insiders, a challenger weighs the power that each insider would add to the new coalition against that insider's own resources, which the new coalition could no longer plunder. When the marginal return to power is high, every insider is worth recruiting, and the most threatening challenge is *enlargement*: outsiders join the entire ruling coalition, because leaving out any insider would cost more in power than it would save in protected resources. The threat of enlargement depends only on the ruling coalition's aggregate power and resources, which ex-

changes among its members leave unchanged, so hierarchy brings no gain in external resilience in this case (Proposition 9).

When the marginal return to power is low, challengers are selective and prefer the insiders whose power is high relative to their resources. If these insiders are strong enough to form a winning coalition with some outsiders, they are the most threatening partners for some groups of outsiders, and the binding threat is *replacement*: the outsiders join these insiders and plunder the remaining insiders. The threat of replacement depends on how power and resources are distributed within the ruling coalition, and hierarchy then strictly increases external resilience (Propositions 5 and 8).[3] A lower marginal return to power also leads to a weakly less powerful ruling coalition (Proposition 7), so more power remains outside the ruling coalition and these insiders can more easily form a winning coalition with outsiders.

The marginal return to power also creates a trade-off between internal and external resilience. A coalition formed by secession has less power than the ruling coalition, whereas a coalition formed by enlargement has more power than the ruling coalition. Hence, when the marginal return to power is high enough for enlargement to be the binding external threat, a further increase in the marginal return to power makes secession less attractive and enlargement more attractive: internal resilience rises and external resilience falls (Proposition 10).

Although the model is static, it can be interpreted as one period of a stationary-bandit economy in the sense of Olson (1993). In each period, the ruling coalition extracts resources from outsiders, and the outsiders rebuild their resources before the next period begins. Plunder can therefore continue indefinitely, and the ruling coalition changes only when a shock occurs. Our results identify the set of shocks that a ruling coalition survives and show how the allocation of power and resources among its members changes this set.

Our main result offers a rationale for a feature that predatory organizations share across very different times and places, namely that rank ties the force that a member commands to the wealth that he holds. Each officer of the Mughal state held a pair of numerical ranks: a personal rank, which fixed his pay, and a cavalry rank, which fixed the contingent that he had to maintain out of that pay. As a rule, the cavalry rank could not exceed the personal rank (Moreland, 1936; Athar Ali, 1966). An Ottoman cavalryman held a revenue grant that obliged him to bring armed retainers to campaign in proportion to the yield of the grant (İnalcık, 1973). More generally, armies tie the pay of their members to rank.

Economic theories of hierarchy explain rank-ordered organizations by incentive provision, coordination, and the organization of knowledge in production (Qian, 1994; Garicano, 2000; Mookherjee, 2013). Our explanation applies instead to organizations whose main activity is extraction. Moreover, it does not rest on the ability of powerful members to take more: the alignment of resources with power matters because it makes the coalition more resilient to outside challenges.

1.1 Relevant Literature

Our model of coalition formation builds on Acemoglu et al. (2008), who study the formation of self-enforcing ruling coalitions in non-democracies under limited commitment; we adapt their

[3]Formally, consider a mis-ordered pair and the set of members whose power-to-resource ratio exceeds a cutoff that lies between the ratios of the two members. The gain is strict if this set of members can form a winning coalition with a proper subset of the outsiders (relevance) and if, at some level of outsider power, it is the unique most threatening partner for outsiders (local decisiveness). Given relevance, a sufficiently low uniform bound on the marginal return to power implies local decisiveness.

axiomatic approach and their protocol (see also Acemoglu et al., 2012). Two features are new. The ruling coalition plunders outsiders and protects the resources of its members, so admitting a member has a cost as well as a benefit. And individuals differ in two dimensions that play opposite roles, so the joint distribution of power and resources within a coalition, and not only the distribution of power, determines its exposure to challenges. The cost of admitting a member recalls the size principle of Riker (1962), under which winning coalitions in zero-sum settings are minimal; here size has two dimensions, and the ruling coalition balances power against protected resources. A related literature studies the allocation of resources by force. In pillage games, a coalition's power can depend on the wealth of its members (Jordan, 2006; MacKenzie et al., 2015); related work studies resource allocation under coercion (Piccione and Rubinstein, 2007) and the robustness of social orders to changes in power relations (Piccione and Razin, 2009). We instead treat power and resources as separate endowments and protect the resources of insiders from plunder. Ray and Vohra (2015) survey broader approaches to coalition formation and stability.

A second literature studies predation and the origins of the state (Olson, 1993; Tilly, 1985; Skaperdas, 1992; Hirshleifer, 1995; Grossman and Kim, 1995; Konrad and Skaperdas, 2012), and recent empirical work studies how states and armed groups arise from, and organize, extraction (Sánchez de la Sierra, 2020; Mayshar et al., 2022; Henn et al., 2025). This literature asks whether and how a predator extracts. We take extraction as given and ask how the internal organization of the extracting coalition affects its survival. North et al. (2009) argue that the dominant coalition of a natural state is held together by rents that tie the interests of elites to its survival, and that the allocation of land and control can be used to rebalance interests within the coalition as power shifts among its members. Our results give a precise sense in which such rebalancing stabilizes a coalition and identify when it matters. In Esteban and Ray (2008), economic inequality within ethnic groups makes them effective in conflict, because the rich supply resources and the poor supply fighters. In our model, by contrast, a member's resources are a liability to any coalition she joins, because they cannot be plundered.

The paper also contributes to the economics of hierarchy. Existing work explains hierarchical organizations by incentive provision and coordination (Qian, 1994; Qian et al., 2006; Mookherjee, 2013) and by knowledge-based specialization (Garicano, 2000; Garicano and Rossi-Hansberg, 2015). We add an explanation based on resilience, which is most relevant where external threats are of first-order importance, as for armies, military aristocracies, and predatory states. Kimya (2026) studies stable status hierarchies and how status rankings align with members' attributes. Our hierarchy instead orders power and resources within a ruling coalition, and our results concern its effect on the coalition's resilience to outsiders.

Finally, the paper contributes to the political economy of non-democracies. A central theme of this literature is that rulers face threats from within, such as coups and elite defection, and from without, such as revolts (Svolik, 2012; Gehlbach et al., 2016; Egorov and Sonin, 2024). One strand examines threats from within the regime, including authoritarian power-sharing, leadership dynamics, and the trade-off between loyalty and competence (Svolik, 2009; Besley and Kudamatsu, 2007; Egorov and Sonin, 2011; Boix and Svolik, 2013; Jia et al., 2015; Zakharov, 2016); another examines threats from outside it (Wintrobe, 1990, 2000; Bueno de Mesquita, 2010). Several papers study how rulers trade the two off: sharing power with potential challengers reduces the risk of rebellion but raises the risk of a coup (Roessler, 2011; Francois et al., 2015; Paine, 2021), and a military strong enough to defeat outsiders may turn against those it protects (Acemoglu et al., 2010; Besley and Robinson, 2010). We study both threats in

one model of coalition formation. In our model, the internal allocation of power and resources affects external resilience without weakening internal stability, while the marginal return to power trades one against the other.

The remainder of the paper is organized as follows. Section 2 introduces the model. Section 3 characterizes the ruling coalition. Section 4 contains our main results on resilience. Section 5 concludes.

2 Environment

The society consists of a finite set of players $N = \{1, 2, \ldots, n\}$, and 2^N denotes the set of all subsets of N. The players are endowed with power and resources, specified by the mappings $p(\cdot) : N \to \mathbb{R}_{++}$ and $x(\cdot) : N \to \mathbb{R}_{++}$. 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 nonempty 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 $P_I := \sum_{i \in I} p_i$ and $X_I := \sum_{i \in I} x_i$. 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 $\beta \in (\frac{1}{2}, 1)$ is a fixed supermajority requirement for power so that the winning coalition can defeat all outsiders. Note that the supermajority requirement applies to power, not to votes. Denote the set of all winning coalitions by $\mathcal{W}$. Throughout, we assume that at least one proper winning coalition exists, so $\mathcal{W} \setminus \{N\} \neq \emptyset$. There is a baseline payoff function $U : N \times (2^N \setminus \{\emptyset\}) \to \mathbb{R}$ that, for any player $i \in N$ and coalition $I \subseteq N$, assigns the payoff $U_i(I)$. When $I \in \mathcal{W}$, this is the payoff if coalition I becomes the "ruling coalition."

A ruling coalition of our model is necessarily a winning coalition. As the key new feature of our setup, a ruling coalition can only plunder outsiders, while the resources of its members are protected. 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 any nonempty coalition $I \subseteq N$, $U_i(I) := x_i + w_i(I)$, where $w_i(\cdot)$ satisfies the following properties:*

1. *(Trade-off) If $I \subsetneq N$ and $i \in I$, we have $w_i(I) = G_i(P_I, X_I) > 0$, where $G_i : (0, P_N) \times \mathbb{R}_{++} \to \mathbb{R}_{++}$ is continuously differentiable, with $\frac{\partial G_i}{\partial P} > 0$ and $\frac{\partial G_i}{\partial X} < 0$ throughout its domain. For every $P \in (0, P_N)$, assume also that $G_i(P, X) \to \infty$ as $X \to 0^+$ and $G_i(P, X) \to 0$ as $X \to \infty$. In particular, G_i satisfies the following conditions at attainable coalition points.*
   (a) *For all $I, I' \subsetneq N$ 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 $I, I' \subsetneq N$ 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 $I \subsetneq N$ and $i \notin I$, then $w_i(I) < 0$.*
3. *For all $i \in N$, $w_i(N) = 0$.*

4. *(Resource feasibility) For every* $I \subsetneq N$, $\sum_{k\in N} w_k(I) = 0$ *and* $w_k(I) \geq -x_k$ *for every* $k \notin I$.

Assumption 1 establishes important primitives of the model. Part 1 introduces the key primitive, the function $G_i(\cdot,\cdot)$. The function $G_i(\cdot,\cdot)$ gives the plundering gain of individual i in any proper coalition of which she is a member, as a function of the coalition's aggregate power and resources.[4] Part 1(a) says that between coalitions with equal power, players prefer the one with fewer internal resources, which permits more external resources for plundering. Meanwhile, between coalitions with equal resources, players prefer the one with larger power (Part 1(b)), as it strengthens the coalition in extracting resources. Part 1 also requires members of a proper coalition to obtain strictly positive gains from plundering outsiders. Together with Part 2, this implies that members of a proper coalition receive payoffs strictly above their initial resources, while outsiders receive payoffs strictly below 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.

Part 4 imposes resource feasibility: insiders' total gains equal outsiders' total losses, and no outsider loses more than her initial resources. Resource feasibility is required only at attainable coalition points. Away from these points, G is only a device for ranking pairs (P, X) and need not be feasible; the limit conditions in Part 1 ensure that every indifference curve is defined at each level of power.

Under Assumption 1, each insider's plundering gains depend on her coalition only through its aggregate 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 plundering gains in coalition I. Assumption 1 immediately yields the following lemma.

Lemma 1 *Under Assumption 1, any player* $i \in N$ *has strictly increasing, continuously differentiable indifference curves over* $(P, X) \in (0, P_N) \times \mathbb{R}_{++}$. *For every plundering gain* $u > 0$, *the indifference curve is the graph* $X = H_i(P; u)$ *on* $P \in (0, P_N)$, *with*

$$H_i'(P;u) = -\frac{\partial G_i/\partial P}{\partial G_i/\partial X}\big(P, H_i(P;u)\big) > 0.$$

The following assumption implies that players rank the coalitions to which they belong in the same way. The resilience analysis in Section 4 relies on this common ranking, but Section 4.4 gives a condition under which the main resilience result extends to heterogeneous preferences.

Assumption 2 *There exist constants* $g(i) > 0$, *one for each* $i \in N$, *and a common continuously differentiable function* $G : (0, P_N) \times \mathbb{R}_{++} \to \mathbb{R}_{++}$ *such that* $G_i(P, X) = g(i)\, G(P, X)$ *for every* $i \in N$ *and every* $(P, X) \in (0, P_N) \times \mathbb{R}_{++}$.

Assumption 2 imposes the factorization $G_i(P, X) = g(i)\, G(P, X)$, which has two components: an idiosyncratic scale $g(i)$ and a common reduced-form function $G(P, X)$. Players therefore share the same indifference curves over coalition power and resources (Figure 1). In

[4]For example, one can view $G_i(P_I, X_I)$ as a combination of a plundering component $F(I) : \{I \subsetneq N : I \neq \emptyset\} \to \mathbb{R}_{++}$ and a share component $\Pi(i, I) : N \times \{I \subsetneq N : I \neq \emptyset\} \to [0, 1]$, 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 $F(I)$.

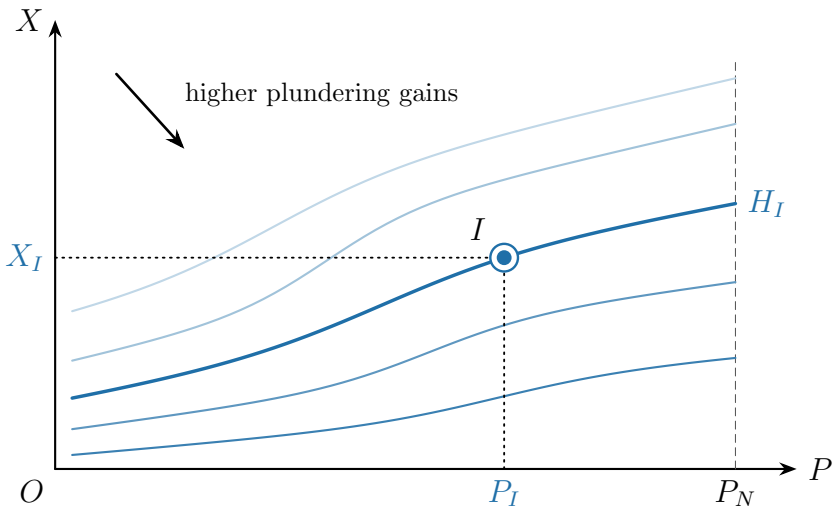

Figure 1: Indifference curves over a coalition's aggregate power P and resources X. Under Assumption 2, all players have the same curves. Each curve is strictly increasing and continuously differentiable on $(0, P_N)$ (Lemma 1); no other restriction on its shape, such as convexity, is imposed. Lower curves give higher plundering gains. The curve through (P_I, X_I) is denoted by H_I.

particular, for any winning 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 $U_j(I) \geq U_j(I')$.

The factorization is imposed on the full domain because the resilience analysis below uses the common indifference curves at all levels of aggregate power and resources, not only at the finitely many attainable coalition points. For a proper coalition I, we write $G(I) := G(P_I, X_I)$. The function $G(I)$ is particularly useful for characterizing the equilibrium coalition and its resilience. For the grand coalition, we set $G(N) := 0$.

Discussion of Assumption 2. Assumption 2 gives a simple consistency condition on insiders' preferences over coalitions. Appendix C shows one environment in which the condition holds. Suppose that insider i's payoff in coalition I can be written as $w_i(I) = \Pi_i(I)F(I)$, where $\Pi_i(I)$ is i's intra-coalition share and $F(I)$ is the coalition's total extractable surplus. For example, $\Pi_i(I)$ may be proportional to $\frac{p_i}{P_I}$, while $F(I)$ depends only on the coalition's aggregate power and on the resources left outside the coalition, $(P_I, X_N - X_I)$. In such an environment, Assumption 2 says that members rank coalitions according to the same aggregate object that determines the coalition's surplus.

Definition 1 *Fix a common function $G(\cdot, \cdot)$ satisfying Assumptions 1–2. For any proper ruling coalition I, denote the indifference curve through (P_I, X_I) by H_I, which is uniquely defined on $P \in (0, P_N)$ by $G(P, H_I(P)) = G(P_I, X_I)$. By Lemma 1, H_I is continuously differentiable and strictly increasing.*

We call the endowment profile *generic* if, for all distinct $I, I' \in \mathcal{W}$, we have $P_I \neq P_{I'}$ or $X_I \neq X_{I'}$. Let $e = (p_1, \ldots, p_n, x_1, \ldots, x_n) \in \mathbb{R}^{2n}_{++}$ denote the endowment profile. For any distinct coalitions I and I', the equalities $P_I = P_{I'}$ and $X_I = X_{I'}$ impose two nontrivial linear restrictions on e. The profiles that satisfy both lie in a proper linear subspace of $\mathbb{R}^{2n}$, which has

measure zero. Since there are finitely many pairs of coalitions, the set of nongeneric endowment profiles is contained in a finite union of measure-zero sets and therefore has measure zero.

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

This assumption implies that players receive strictly different payoffs from different winning coalitions that include them. We use it to establish that the baseline model has a unique ruling coalition. Uniqueness provides a basis for the resilience analysis in Section 4, which studies how the ruling coalition responds to exogenous shocks to players' power and resources. The resilience results themselves allow ties among coalitions other than the ruling coalition.

3 Preliminary Analysis of the Coalition Formation Game

This section establishes existence and uniqueness of the coalitional equilibrium.

3.1 Axiomatic analysis

We begin with an axiomatic analysis. As in Acemoglu et al. (2008) and Acemoglu et al. (2012), 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 $I \in \mathcal{W}$, $\phi(I) \neq \emptyset$, $\emptyset \notin \phi(I)$, and $N \notin \phi(I)$.*

Axiom 2 (Supermajority 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 $I'' \in \mathcal{W}$, we have $I'' \notin \phi(I)$ if and only if $G(I'') < G(I')$.*

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 that ϕ maps any initial winning coalition to a nonempty set of non-trivial ruling coalitions, excluding both the empty set $\emptyset$ and the grand coalition N. Axiom 2 requires every 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 weakly higher $G(\cdot)$ must be selected. Proposition 1 establishes that these axioms identify a unique mapping under Assumptions 1–2, and that the correspondence is single-valued under Assumptions 1–3.

[5] At a generic endowment profile, a tie can be removed by an arbitrarily small perturbation of G. A small downward C^1 perturbation supported near one of the tied winning points breaks the tie while preserving resource feasibility, and the strict inequalities among the other winning coalitions persist when the perturbation is sufficiently small. Appendix C gives a parametric family in which ties arise for at most finitely many parameter values.

Axiom 3 is deliberately strong and selects directly the winning coalitions with the highest G, which gives the simple characterization on which the resilience analysis in Section 4 rests. Unlike in Acemoglu et al. (2008), no separate self-enforcement axiom is needed, because under Assumption 2 no winning sub-coalition of such a coalition gives its members strictly higher plundering gains. Section 4.4 relaxes the common ranking of Assumption 2 and gives a condition under which the main resilience result still holds.

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

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

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

Proposition 1 shows that the ruling coalition is the winning coalition that maximizes its members' plundering gains, 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

The extensive form adapts the coalition-formation protocol of Acemoglu et al. (2008) to plundering. 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 Assumptions 1–2, and $\beta \in (\frac{1}{2}, 1)$ is the supermajority requirement. The game starts with the initial winning coalition $I_0 \in \mathcal{W}$, and the steps are as follows:

1. Let $q \in \{1, \ldots, |I_0|\}$ denote the round of agenda setting and voting. In round $q = 1$, Nature randomly picks an agenda setter a_1 from the initial winning coalition I_0.

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 in which the members of I_q vote 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 in round $q + 1$. 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). We first describe the players' strategies. A pure strategy of any player $i \in I_0$ is a pair of functions $\sigma_i = (v_i, \mathcal{P}_i)$ 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 at any point in 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 ϕ 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 1–2 hold, that ϕ 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 follows from the axiomatic characterization in Proposition 1, where $\phi(I_0) = \arg\max_{W \in \mathcal{W}} G(W)$ for any $I_0 \in \mathcal{W}$. Consider any $I \in \phi(I_0)$. When an agenda setter in $I \cap I_0$ proposes I, each member of I weakly prefers it to the continuation after rejection. Before such an agenda setter is reached, each winning alternative includes a member of I who can reject it in anticipation of I. If I is the unique maximizer, its members strictly prefer it to every other outcome, so every SPE yields I. Assumption 3 ensures this uniqueness in the baseline game.

Remark: potential ruling coalitions. We now define the set of *potential ruling coalitions*, which contains the equilibrium ruling coalition and is a useful tool for understanding coalitional resilience.

Definition 2 (Potential ruling coalitions) *A winning coalition I is a* potential ruling coalition *if no winning coalition has strictly more power and strictly fewer resources than I. The set of potential ruling coalitions is*

$$Z := \{I \in \mathcal{W} \mid \text{there is no } I' \in \mathcal{W} \text{ with } P_{I'} > P_I \text{ and } X_{I'} < X_I\}. \tag{1}$$

Figure 2 illustrates the potential ruling coalitions (blue dots). The ruling coalition in Proposition 2 maximizes G among winning coalitions, and G increases with power and decreases with resources. The ruling coalition must therefore be a potential ruling coalition—a simple observation that will prove useful in what follows.

4 Coalitional resilience

This section contains the main results of the paper. We first characterize the unique ruling coalition (Proposition 3), define internal and external resilience, and relate external resilience to the probability that the incumbent survives shocks to outsiders. We then show that admissible exchanges lead to a hierarchical allocation and weakly increase external resilience (Proposition 4

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

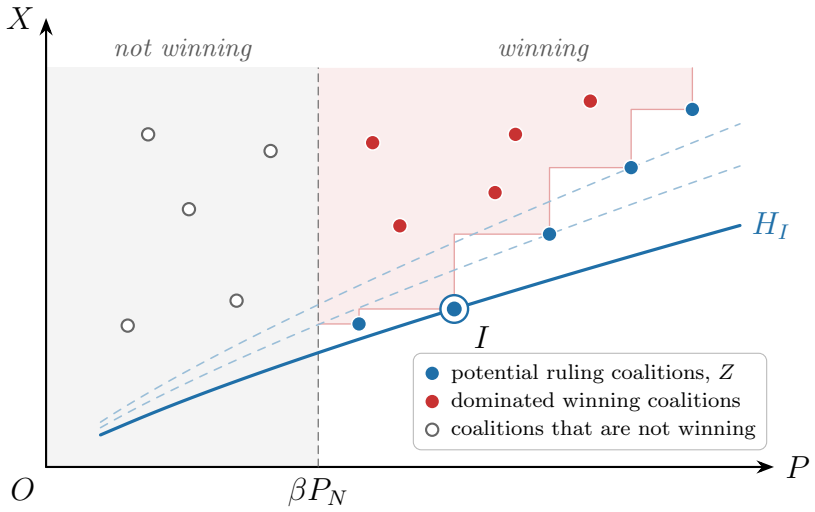

Figure 2: The set of potential ruling coalitions Z and the unique ruling coalition I. Each dot is a coalition, placed at its aggregate power and resources. A winning coalition is a potential ruling coalition (blue) if no winning coalition has strictly more power and strictly fewer resources; the other winning coalitions (red) lie in the shaded area. The ruling coalition I is the potential ruling coalition on the lowest indifference curve, H_I, where plundering gains are highest. The dashed curves are indifference curves with lower plundering gains.

and Corollary 1). We give a condition under which the increase is strict (Proposition 5) and extend the result for a single exchange to heterogeneous preferences (Proposition 6). Finally, we show how the gain from hierarchy depends on the marginal return to power (Propositions 7, 8, and 9) and derive a trade-off between internal and external resilience (Proposition 10).

4.1 Internal and external resilience

A winning coalition is the unique ruling coalition if and only if it is the unique maximizer of G among winning coalitions (Proposition 1). This subsection divides the other winning coalitions into two kinds of threat—those from its own winning sub-coalitions and those from winning coalitions that include outsiders—and shows that it is enough to check a small set of each kind. We use this division to define two notions of resilience, internal and external. The distinction corresponds to the two threats emphasized in the literature on non-democratic regimes, those from regime insiders and those from outsiders (Svolik, 2012; Meng, 2020; Paine, 2021; Egorov and Sonin, 2024).

We begin by defining the best sub-coalitions of a set of players.

Definition 3 (Best sub-coalitions) *Let $I \subseteq N$ be a nonempty set of players. A nonempty sub-coalition $A \subseteq I$ is a* best sub-coalition *of I if there is no sub-coalition $A' \subseteq I$ with*

$$P_{A'} > P_A \quad \text{and} \quad X_{A'} < X_A. \tag{2}$$

We denote the set of best sub-coalitions of I by $\mathcal{A}_I$.

For any $I \subseteq N$, $\mathcal{A}_I$ consists of all nonempty subsets of I that are undominated in power and resources: no other subset of I has both strictly higher power and strictly lower resources.

Equation (2) is analogous to Equation (1) in Definition 2 for potential ruling coalitions, but restricts attention to sub-coalitions of a coalition I. We can now state Proposition 3, which characterizes the two types of threats to the ruling coalition. In particular, given a ruling coalition I, denote by $A^{\text{ins}} \in (\mathcal{A}_I \setminus \{I\}) \cap \mathcal{W}$ a proper best sub-coalition of I that is also a winning coalition. A coalition A^{ins} is therefore an internal threat. Any best sub-coalition of I, including I itself, can also form an alliance with outsiders to challenge the ruling coalition I, and these outsiders should also form a best sub-coalition of all outsiders. We denote a coalition of such outsiders by $A^{\text{ext}} \in \mathcal{A}_{N\setminus I}$. A ruling coalition only needs to defeat any internal threat A^{ins} and every winning alliance between a best sub-coalition of I and the most threatening outsiders A^{ext}, as established by the following proposition.

Proposition 3 (Internal and external threats) *Consider the game* Γ *and suppose that Assumptions 1–2 hold. A coalition* I *is the unique ruling coalition, that is,* $\phi(I_0) = \{I\}$*, if and only if* I *is winning and the following two conditions hold.*

(i) No profitable internal secession. $G(I) > G(A^{\text{ins}})$ *for every proper winning best sub-coalition* A^{ins} *of* I*.*

(ii) No profitable external deviation. $G(I) > G(A^{\text{ins}} \cup A^{\text{ext}})$ *for every* $A^{\text{ins}} \in \mathcal{A}_I$ *and every* $A^{\text{ext}} \in \mathcal{A}_{N\setminus I}$ *such that* $A^{\text{ins}} \cup A^{\text{ext}}$ *is winning.*

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 and best sub-coalitions of outsiders.

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). Our analysis of internal resilience is based on the set of such points.

Definition 4 (Internal safe area and internal resilience) *The* internal safe area *of a ruling coalition* I *is*

$$\mathcal{S}_I^{\text{int}} := \big\{(P, X) \in (0, P_N) \times \mathbb{R}_{++} \;\big|\; P < \beta P_N \;\; \text{or} \;\; X > H_I(P)\big\}. \tag{3}$$

A ruling coalition I *is* weakly more internally resilient *than a ruling coalition* J *if* $\mathcal{S}_J^{\text{int}} \subseteq \mathcal{S}_I^{\text{int}}$*, and* strictly more internally resilient *if* $\mathcal{S}_J^{\text{int}} \subsetneq \mathcal{S}_I^{\text{int}}$*. The two coalitions are* equally internally resilient *if* $\mathcal{S}_I^{\text{int}} = \mathcal{S}_J^{\text{int}}$*.*

An *internal shock* redistributes power and resources among the members of I. It leaves unchanged the aggregate power and resources of I, the endowments of outsiders, and the function G. We consider only internal shocks that preserve resource feasibility (Assumption 1). We compare internal resilience to such shocks by comparing safe areas. The resulting ranking holds whatever the probability distribution of internal shocks: a coalition with a larger internal safe area withstands every internal shock that it would withstand with a smaller safe area.

For any coalition I to be the ruling coalition, all its proper winning best sub-coalitions must lie in $\mathcal{S}^{\text{int}}$. This guarantees that all members of I prefer I to any proper winning best sub-coalition of I, satisfying Condition (i) of Proposition 3. Non-winning sub-coalitions are automatically safe. We write $\mathcal{S}^{\text{int}}$ for $\mathcal{S}_I^{\text{int}}$ when there is no confusion. Figure 3 illustrates the

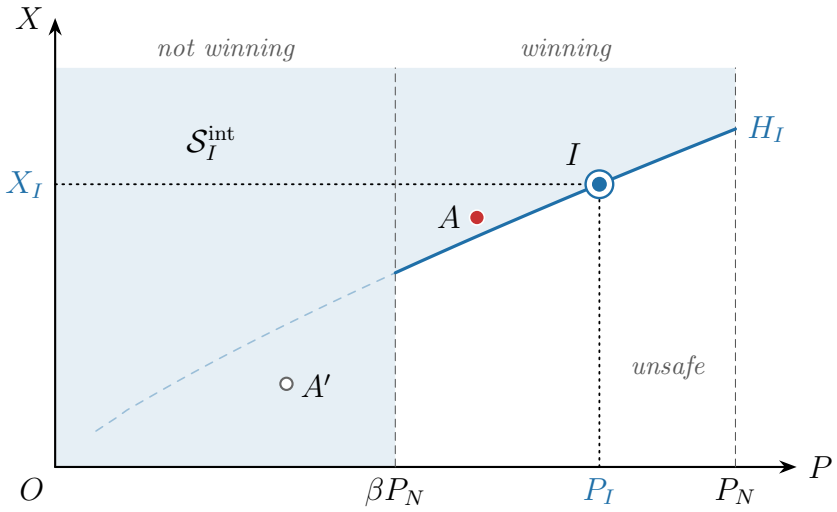

Figure 3: The internal safe area $\mathcal{S}_I^{\text{int}}$ (shaded). It consists of the points to the left of βP_N, where a sub-coalition is not winning, and the points above the indifference curve H_I through (P_I, X_I). The sub-coalition A is winning and lies above H_I, so its members prefer I. The sub-coalition A' lies below H_I but is not winning. Both are safe.

internal safe area $\mathcal{S}^{\text{int}}$ for a coalition I. This set plays a central role in our subsequent analysis of internal resilience.

We now turn to "external" threats. The grand-coalition deviation is automatically unprofitable because $G(N) = 0 < G(I)$, so the geometry below concerns proper deviations. For any best sub-coalition of outsiders A^{ext} and any best sub-coalition of insiders A^{ins} whose union is a proper winning coalition, Condition (ii) of Proposition 3 requires $G(I) > G(A^{\text{ins}} \cup A^{\text{ext}})$, 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

$$X_{A^{\text{ext}}} > H_I(P_{A^{\text{ins}}} + P_{A^{\text{ext}}}) - X_{A^{\text{ins}}}. \tag{4}$$

For any group of outsiders A^{ext} that satisfies Condition (4), the group of outsiders cannot threaten the ruling coalition together with A^{ins}. We need to make sure that Condition (4) is satisfied for every best sub-coalition of the ruling coalition with which A^{ext} forms a proper winning coalition. This motivates the following definition of the "external safe area" for a ruling coalition I; Figure 4 illustrates its construction.

Definition 5 (External safe area and external resilience) *Let I be a ruling coalition, and let $O := N \setminus I$ denote the set of outsiders. The set of possible aggregates of an outsider coalition is $\mathcal{D}_O := \big([0, P_O) \times \mathbb{R}_{++}\big) \cup \{(P_O, X_O)\}$. For a nonempty sub-coalition $A \subseteq I$, the* translated boundary *of A at outsider power $P \in [0, P_O]$ is*

$$H_I^A(P) := \begin{cases} H_I(P + P_A) - X_A, & \text{if } \beta P_N \le P + P_A < P_N, \\ -\infty, & \text{otherwise.} \end{cases} \tag{5}$$

The translated boundary of A is active *at P if $H_I^A(P)$ is finite, that is, if A and an outsider coalition with power P form a proper winning coalition.*

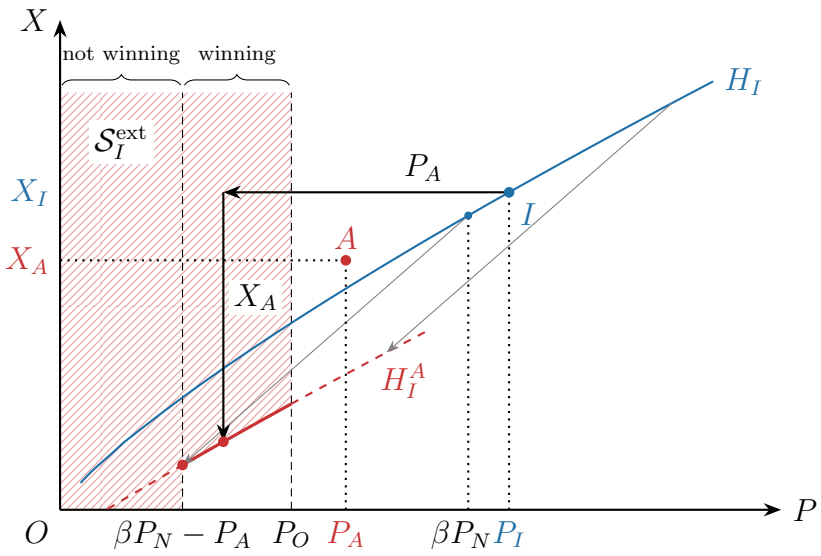

Figure 4: Construction of the external safe area. A best sub-coalition A of I has aggregate power and resources (P_A, X_A). Shifting the indifference curve H_I (blue) left by P_A and down by X_A gives the curve $H_I(P+P_A)-X_A$ (red); the arrows show the shift. The translated boundary H_I^A is the part of this curve (thick) at which A and outsiders with power P form a winning coalition, that is, $P \geq \beta P_N - P_A$, up to the outsiders' total power P_O. The shaded area is the external safe area $\mathcal{S}_I^{\text{ext}}$ determined by A alone: the points at which outsiders cannot form a profitable deviation with A. With several best sub-coalitions, $\mathcal{S}_I^{\text{ext}}$ is the intersection of these areas.

The upper envelope *of the translated boundaries of the best sub-coalitions of I is*

$$\overline{H}_I(P) := \max_{A \in \mathcal{A}_I} H_I^A(P), \tag{6}$$

and the external safe area *of I is*

$$\mathcal{S}_I^{\text{ext}} := \big\{(P,X) \in \mathcal{D}_O \;\big|\; X > \overline{H}_I(P)\big\}. \tag{7}$$

A ruling coalition I is weakly more externally resilient *than a ruling coalition J if $\mathcal{S}_J^{\text{ext}} \subseteq \mathcal{S}_I^{\text{ext}}$, and* strictly more externally resilient *if $\mathcal{S}_J^{\text{ext}} \subsetneq \mathcal{S}_I^{\text{ext}}$. The two coalitions are* equally externally resilient *if $\mathcal{S}_I^{\text{ext}} = \mathcal{S}_J^{\text{ext}}$.*

What is the intuition behind Definition 5 and the notion of external safety? First, since I is winning and $\beta > \frac{1}{2}$, every outsider coalition has power at most $P_O \leq (1-\beta)P_N < \beta P_N$. Thus any winning external deviation must include at least one insider.

Second, fix a best sub-coalition of insiders A. If $P + P_A < \beta P_N$, then A cannot combine with outsiders of power P to form a winning coalition. If $P + P_A = P_N$, the coalition is the grand coalition and generates no plunder. For a proper winning coalition, a point on the translated boundary $H_I^A(P)$ makes the insiders in A indifferent between the current ruling coalition I and the alternative. Above the boundary, those insiders strictly prefer I. The external safe area $\mathcal{S}_I^{\text{ext}}$ consists of the points above the upper envelope of the translated boundaries over all $A \in \mathcal{A}_I$.

External shocks and coalition survival. The external safe area $\mathcal{S}_I^{\text{ext}}$ is a condition on the aggregate power and resources of an individual outsider coalition: a coalition $B \subseteq O$ is safe if $(P_B, X_B) \in \mathcal{S}_I^{\text{ext}}$. An outsider shock, by contrast, may change the endowments of several outsiders simultaneously, and hence the aggregates (P_B, X_B) of several outsider coalitions.

Fix the incumbent I, its internal allocation of power and resources, and the outsiders' aggregate power and resources (P_O, X_O). A *state* is a profile of outsider endowments $z = \big((p_k(z), x_k(z))\big)_{k\in O}$, with $p_k(z) > 0$ and $x_k(z) > 0$ for every $k \in O$, such that $\sum_{k\in O} p_k(z) = P_O$ and $\sum_{k\in O} x_k(z) = X_O$. Let $\mathcal{Z}_O$ denote the set of states that satisfy resource feasibility under Assumption 1, with G held fixed. In every state, the game starts from I. Neither the power of I nor the winning threshold βP_N depends on z, so the incumbent remains a winning coalition and its value $G(I)$ is the same in every state. The incumbent *survives* at z if I is the unique ruling coalition at z. Let $E_I \subseteq \mathcal{Z}_O$ denote the set of states at which the incumbent survives, and let $\mathcal{A}_O(z)$ denote the set of best outsider sub-coalitions at z.

Suppose that I is internally stable, that is, Condition (i) of Proposition 3 holds. Internal stability does not depend on z, because the internal allocation of I and the winning threshold are the same in every state. Moreover, Definition 5 applies to every outsider coalition B in every state: if $B \subsetneq O$, then $P_B(z) < P_O$, and if $B = O$, then $\big(P_B(z), X_B(z)\big) = (P_O, X_O)$. Proposition 3 therefore implies

$$z \in E_I \quad \text{if and only if} \quad \big(P_B(z), X_B(z)\big) \in \mathcal{S}_I^{\text{ext}} \quad \text{for every } B \in \mathcal{A}_O(z). \tag{8}$$

That is, the incumbent survives at z if and only if every best outsider sub-coalition at z is safe.

Consider two internally stable allocations, I and I', of the same members with the same aggregate power and resources. Restrict $\mathcal{Z}_O$ to the states at which resource feasibility holds under both allocations. If $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{I'}^{\text{ext}}$, then Equation (8) implies $E_I \subseteq E_{I'}$: the incumbent survives under I' in every state in which it survives under I. Hence $\mu(E_I) \leq \mu(E_{I'})$ for any probability distribution μ on $\mathcal{Z}_O$. That is, the incumbent is at least as likely to survive under I' as under I. This result places no restriction on μ; in particular, the endowments of different outsiders may be arbitrarily dependent.

The inequality is strict if and only if $\mu(E_{I'} \setminus E_I) > 0$, that is, if and only if the incumbent survives under I' but not under I with positive probability. A strict inclusion $\mathcal{S}_I^{\text{ext}} \subsetneq \mathcal{S}_{I'}^{\text{ext}}$ does not by itself guarantee this condition. There are three reasons. First, the additional safe points may not correspond to any outsider coalition in any state in $\mathcal{Z}_O$. For example, if O consists of a single player, then $\mathcal{Z}_O$ contains a single state. Second, in a state where one outsider coalition is safe only under I', another outsider coalition may be unsafe under both allocations, so the incumbent does not survive under either. Third, even if such states exist, μ may assign them zero probability.

4.2 Which ruling coalitions are more externally resilient?

Consider a ruling coalition I and two members $i, j \in I$. We say that i and j are *weakly misordered* if $p_i \geq p_j$ and $x_i \leq x_j$, with at least one inequality strict. An *exchange* between such members transfers power from i to j, resources from j to i, or both, without reversing either ordering, and leaves the endowments of all other players unchanged (Figure 5). An exchange therefore leaves the aggregate power and resources of I unchanged.

An exchange does, however, change the aggregates of every coalition that contains exactly one of i and j. Under Assumption 2, resource feasibility requires $G(P_C, X_C) \sum_{k\in C} g(k) \leq$

$X_N - X_C$ for every proper coalition C, so feasibility before an exchange does not ensure feasibility after it. We therefore call an exchange *admissible* if Assumptions 1–2 hold both before and after it, with the same function G. A sequence of exchanges is admissible if every exchange in it is admissible. Intuitively, admissibility requires that no coalition plunders more than the resources outside it. This requirement is met whenever total plunder is a share of those resources.

The following proposition shows that an admissible exchange weakly increases the external resilience of the ruling coalition. The result holds for *every* function G that satisfies Assumptions 1–2.

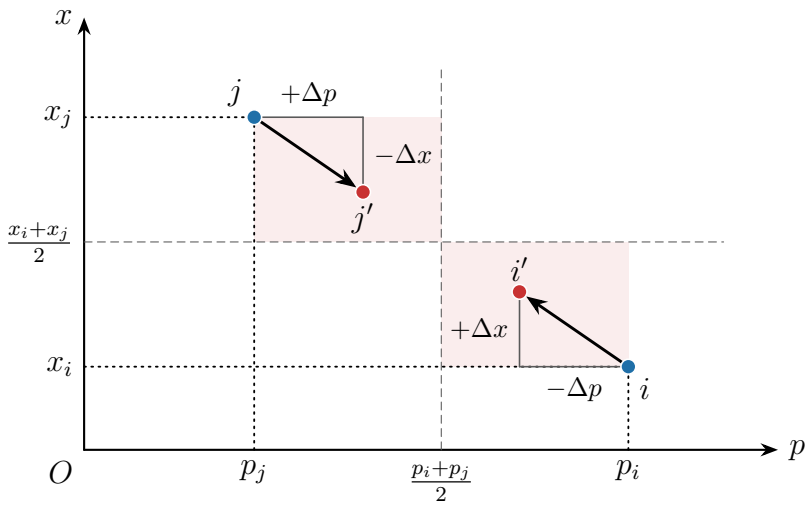

Figure 5: An exchange between players i and j, from blue to red. Player i has more power and fewer resources than player j. She gives up power Δp and receives resources Δx, and j does the reverse. The bounds $\Delta p \leq \frac{p_i - p_j}{2}$ and $\Delta x \leq \frac{x_j - x_i}{2}$ keep i' and j' in the shaded rectangles, so neither ordering is reversed. An exchange may also change only power or only resources.

Proposition 4 *Suppose that I is the unique ruling coalition of the game Γ, and let $i, j \in I$ be weakly mis-ordered. Consider an exchange $(\Delta p, \Delta x) \neq (0, 0)$ between i and j with $0 \leq \Delta p \leq \frac{p_i - p_j}{2}$ and $0 \leq \Delta x \leq \frac{x_j - x_i}{2}$, and let $I' := (I \setminus \{i, j\}) \cup \{i', j'\}$ denote the coalition after the exchange, where*

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

If the exchange is admissible, then I' is the unique ruling coalition after the exchange, and I' is weakly more externally resilient than I:

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

After an exchange, the game starts from the same set of members, which remains a winning coalition because its power is unchanged. The proof of Proposition 4 shows that Condition (i) of Proposition 3 remains satisfied. Because the external safe area weakly expands, Condition (ii) remains satisfied as well. Hence the same set of members remains the unique ruling coalition after an admissible exchange.

Proposition 4 has a direct implication for the survival of the ruling coalition, which we call the incumbent. By Equation (8), the inclusion $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{I'}^{\text{ext}}$ implies $E_I \subseteq E_{I'}$. Hence

$\mu(E_I) \leq \mu(E_{I'})$ for any probability distribution μ on $\mathcal{Z}_O$: an admissible exchange weakly increases the probability that the incumbent survives.

Proposition 4 is our main result on coalitional resilience. The proof is in Appendix A and proceeds in three steps, which we now outline.

Step 1. The exchange has two effects on the best sub-coalitions of I. (i) Every best sub-coalition that contains i but not j loses power, gains resources, or both, and therefore becomes less threatening: it yields lower plundering gains, alone or together with outsiders. (ii) The set of best sub-coalitions may itself change: some sub-coalitions may become best sub-coalitions, and others may cease to be.

Step 2. Neither effect reduces external resilience. Consider effect (i). The exchange leaves the aggregate power and resources (P_I, X_I) of the ruling coalition unchanged, so the indifference curve $H_I(\cdot)$ is also unchanged. Let A be a best sub-coalition that contains i but not j. Its translated boundary is active only where $P + P_A \geq \beta P_N$; at lower values of P, A cannot form a winning coalition with outsiders. Wherever the boundary remains active after the exchange, it is strictly lower, because A has less power, more resources, or both. The translated boundaries of these sub-coalitions therefore impose weaker constraints on external safety.

For effect (ii), we show that every best sub-coalition after the exchange is weakly less threatening than some best sub-coalition before it. Let A' be a best sub-coalition of I'. We first construct a sub-coalition $A \subseteq I$ with weakly more power and weakly fewer resources than A'. If A' contains both i' and j', or neither, let A consist of the same members before the exchange. If A' contains i' but not j', replace i' by i. If A' contains j' but not i', replace j' by i. Next, if $A \in \mathcal{A}_I$, set $B = A$. Otherwise, replace A by a sub-coalition of I with strictly more power and strictly fewer resources, and repeat until the result belongs to $\mathcal{A}_I$. Because I is finite, this procedure ends at some $B \in \mathcal{A}_I$, and by construction $P_B \geq P_{A'}$ and $X_B \leq X_{A'}$.

Where the translated boundary of A' is active, that is, where $\beta P_N \leq P + P_{A'} < P_N$, it therefore lies weakly below the translated boundary of B. Hence the upper envelope of the active translated boundaries does not rise after the exchange, and the external safe area does not shrink.

Step 3. The exchange does not create a profitable internal secession, that is, a proper winning sub-coalition of I' whose members would gain from ruling without the other members. Before the exchange, I satisfies Condition (i) of Proposition 3: $G(I) > G(A)$ for every proper winning best sub-coalition A of I. By the construction in Step 2, every proper winning best sub-coalition A' of I' has weakly less power and weakly more resources than some proper winning best sub-coalition B of I. Therefore $G(A') \leq G(B) < G(I) = G(I')$, where the equality holds because the exchange leaves (P_I, X_I) unchanged. Hence I' also satisfies Condition (i).

A new justification for hierarchy. Proposition 4 shows that one admissible exchange makes the ruling coalition weakly more externally resilient. We now show that a sequence of exchanges leads to a particular internal allocation, which we call a hierarchy.

Definition 6 (Hierarchy) *An internal allocation of a coalition is* hierarchical *if its members can be partitioned into* ranks $R_1, \ldots, R_m$ *such that:*

(i) all members of rank R_r have the same power and resources, denoted by (p^r, x^r); and

(ii) for every $r < s$, we have $p^r \geq p^s$, $x^r \geq x^s$, and $(p^r, x^r) \neq (p^s, x^s)$.

A rank may consist of a single member. We say that rank R_r is *higher* than rank R_s if $r < s$. Each member of a higher rank thus has weakly more power and weakly more resources than each member of a lower rank, and strictly more of at least one. Equivalently, an internal allocation is hierarchical if and only if it contains no pair of members i, j with $p_i > p_j$ and $x_i < x_j$. We call such a pair *mis-ordered*. Figure 6 illustrates a hierarchical allocation.

A hierarchy in this sense is an alignment of power and resources, not a prescription of inequality. An allocation in which all members have the same power and resources is hierarchical, with a single rank ($m = 1$), and so is any allocation in which no member has more power and fewer resources than another. What the definition excludes is a member who is strong but poor relative to another member.

Corollary 1 (Hierarchy through exchanges) *Suppose that I is the unique ruling coalition. Then there exists a finite sequence of exchanges as in Proposition 4 that turns the internal allocation of I into a hierarchical allocation, without changing the aggregate power and resources of I. If the sequence is admissible, then after every exchange the coalition is the unique ruling coalition and is weakly more externally resilient than before that exchange. In particular, the hierarchical allocation is weakly more externally resilient than I.*

We now provide an algorithm that reaches a hierarchy in finitely many exchanges. This is important because an exchange can create a new mis-ordered pair: one of the two members may become mis-ordered with a third member. Exchanges made in an arbitrary order are therefore not guaranteed to reach a hierarchy in finitely many steps.

The sequence is constructed as follows. A *block* is a set of members who have the same power and resources. Begin with each member as a separate block, and order the blocks from highest to lowest power per member; if two blocks have the same power per member, place the block with more resources per member first. Two adjacent blocks are mis-ordered if the block with more power per member has strictly fewer resources per member. Whenever two adjacent blocks are mis-ordered, we *pool* them: we give every member of the two blocks the average power and the average resources of all their members.

Pooling two blocks can be carried out by finitely many exchanges as in Proposition 4. Let (p^*, x^*) denote the average power and resources of the members of the two blocks. Because the two blocks are mis-ordered, their endowments lie on a line $x = c - bp$ with $b > 0$, and (p^*, x^*) lies on the same line. Choose a member i with $p_i > p^*$ and a member j with $p_j < p^*$, and let them exchange $\Delta p = \min\{p_i - p^*,\, p^* - p_j\}$ and $\Delta x = b\,\Delta p$. After this exchange, both members remain on the line, and at least one of them has the endowment (p^*, x^*). The exchange satisfies both inequalities of Proposition 4: $\Delta p \leq \frac{p_i - p_j}{2}$ because the minimum of two numbers is at most their average, and $\Delta x \leq \frac{x_j - x_i}{2}$ because $x_j - x_i = b(p_i - p_j)$.

Repeat this step with another such pair. The average power of the members remains p^* after every exchange, so such a pair exists until every member has the endowment (p^*, x^*). Each exchange brings at least one more member to (p^*, x^*), so the pooling ends after finitely many exchanges.

Each pooling replaces two blocks by one, so at most $|I| - 1$ pooling steps occur. The power per member of the pooled block lies strictly between those of the two original blocks, so the pooled block takes their place in the ordering and the ordering rule continues to hold. The algorithm ends when no two adjacent blocks are mis-ordered. At that point, power per member and resources per member both weakly decrease along the ordering. Combining blocks

with identical endowments then gives the ranks of Definition 6.[7] If the sequence is admissible, Proposition 4 can be applied to the exchanges one after another: after each exchange, the coalition is again the unique ruling coalition, so the proposition applies to the next exchange as well. Appendix C gives a numerical example of an admissible sequence of exchanges that leads to a hierarchical allocation.

Corollary 1 thus provides a resilience rationale for aligning resources with power across the ranks of a plundering coalition, such as those of stable oligarchies, Weberian bureaucracies, and armies.[8]

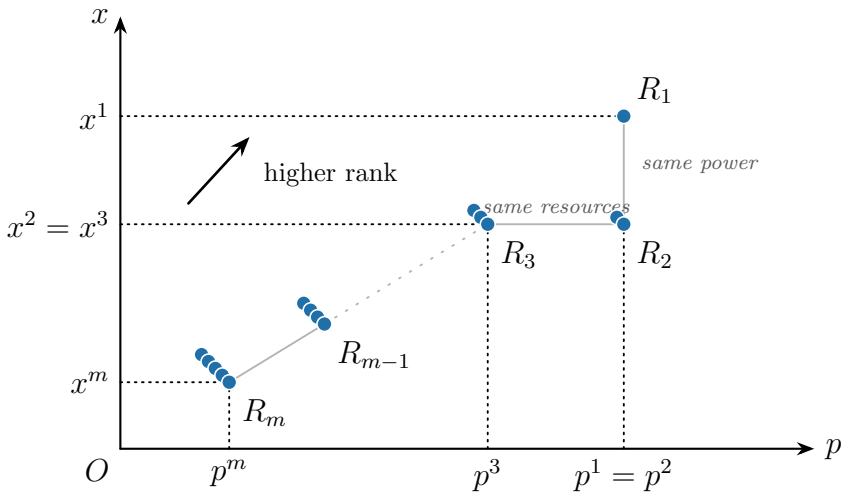

Figure 6: A hierarchical allocation with m ranks. Each stack of dots is a rank R_r, whose members all have the same power and resources (p^r, x^r); a rank may consist of a single member. A higher rank has weakly more power and weakly more resources than a lower rank, and strictly more of at least one. Here R_2 has more power than R_3 and the same resources, and R_1 has more resources than R_2 and the same power.

Corollary 1 compares a hierarchical allocation only with the allocations that lead to it through exchanges. It does not rank two hierarchical allocations against each other; such a ranking requires additional structure on G. For example, consider a ruling coalition I with two members and two hierarchical internal allocations, $\{i, j\}$ and $\{i', j'\}$, with the same aggregate power and resources (Figure 7), that is, $p_i + p_j = p_{i'} + p_{j'} = P_I$ and $x_i + x_j = x_{i'} + x_{j'} = X_I$. Without further restrictions on G, these two allocations cannot be ranked by external resilience.

Dynamic implications. The model is static, but plunder itself changes the distribution of resources. With power held fixed, plunder reduces the resources of outsiders, so a coalition that includes them holds fewer resources and yields higher plundering gains. Plundered outsiders

[7]The pooling procedure does not, in general, equalize all members. Suppose that I contains a mis-ordered pair. Then the procedure ends with a single rank if and only if, for every $k < |I|$, the average resources of the k most powerful members are strictly below the average resources of all members; see Condition (20) in Appendix A.

[8]This result has a natural design interpretation. Suppose that a designer chooses the internal allocation ex ante, before the endowments of outsiders are realized, to maximize the probability that the coalition survives. Then, for any distribution of shocks, it is without loss of generality to restrict attention to hierarchical allocations; see Corollary 3 in Appendix A.

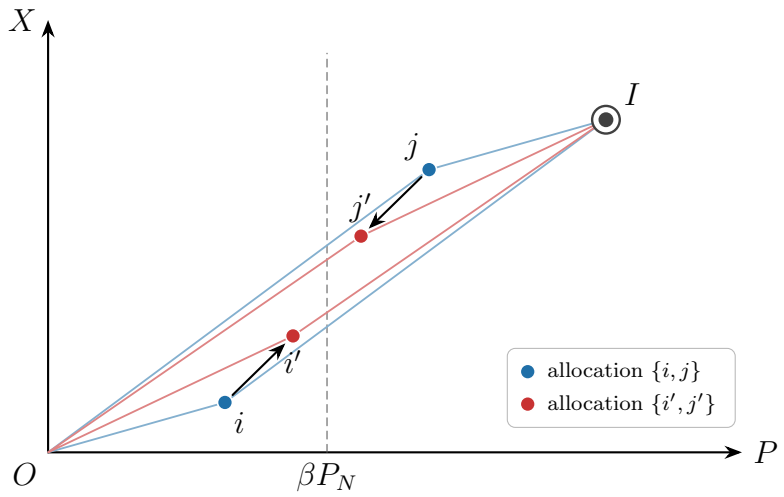

Figure 7: Two hierarchical allocations that cannot be ranked by external resilience. The allocations $\{i, j\}$ (blue) and $\{i', j'\}$ (red) have the same aggregate I: along each colored outline, the two members' power and resources add up to I. Relative to $\{i, j\}$, the second allocation gives the higher-ranked member less power and fewer resources, and the lower-ranked member more of both. Without further restrictions on G, neither allocation is more externally resilient than the other.

may thus become more attractive coalition partners in later periods and may destabilize the coalition that plundered them.

The game and the resilience analysis can be interpreted as one period of a repeated *stationary-bandit* environment (Olson, 1993). In each period, the ruling coalition extracts and consumes the resources of outsiders. Suppose that it preserves their productive capacity, for instance by not destroying land or eliminating the labor force, and that outsiders rebuild their initial resources through production, for example in agriculture, before the next period begins. The next period then starts from the same distribution of resources $x(\cdot)$, and the same coalition-formation game is played again. In the absence of exogenous shocks, plunder can therefore continue indefinitely. In such an environment, changes in the ruling coalition arise from shocks, and our resilience analysis identifies the shocks that the ruling coalition survives. Appendix B discusses this interpretation further.

Proposition 4 shows that an admissible exchange *weakly* increases the external resilience of the ruling coalition. The following subsections give conditions under which the increase is *strict*, and conditions under which external resilience is unchanged. These conditions depend on the slope of the indifference curve $H_I(\cdot)$, which measures the marginal return to power.

4.3 Strict gains in external resilience from hierarchy

This subsection gives a sufficient condition under which an admissible sequence of exchanges that leads to a hierarchical allocation *strictly* increases external resilience. Section 4.5 relates this condition to a low marginal return to power. We begin by defining cutoff coalitions, the sub-coalitions on which the condition is imposed, and show in a lemma that they are best sub-coalitions.

For a coalition I and a scalar $\lambda > 0$, the *cutoff coalition* is $A_\lambda := \{k \in I : p_k - \lambda x_k > 0\}$,

the set of members of I whose power-to-resource ratio p_k/x_k exceeds λ.

Lemma 2 (Cutoff coalitions are best sub-coalitions) *Let I be a coalition.*

(i) For every $\lambda > 0$ such that $A_\lambda \neq \emptyset$, we have $A_\lambda \in \mathcal{A}_I$.

(ii) If $i, j \in I$ are mis-ordered, with $p_i > p_j$ and $x_i < x_j$, then $i \in A_\lambda$ and $j \notin A_\lambda$ for every $\lambda \in \left(\frac{p_j}{x_j}, \frac{p_i}{x_i}\right)$.

The proof of Lemma 2 is in Appendix A. The idea is as follows. For part (i), note that $P_A - \lambda X_A = \sum_{k \in A}(p_k - \lambda x_k)$ for every sub-coalition $A \subseteq I$. The cutoff coalition maximizes this sum, because it contains exactly the members k for whom $p_k - \lambda x_k$ is positive. A sub-coalition with strictly more power and strictly fewer resources than A_λ would attain a strictly higher value of $P_A - \lambda X_A$, so no such sub-coalition exists. Hence A_λ is a best sub-coalition of I whenever it is nonempty. For part (ii), a mis-ordered pair satisfies $\frac{p_i}{x_i} > \frac{p_j}{x_j}$, so for any λ between the two ratios, A_λ contains i but not j.

A cutoff coalition that contains i but not j is important for the following reason. An exchange between i and j lowers the translated boundary of every best sub-coalition that contains i but not j. The external safe area strictly expands if one of these boundaries alone determines the upper envelope of the translated boundaries at some level of outsider power. Our sufficient condition requires this of the cutoff coalition A_λ, a property we call local decisiveness. Local decisiveness implies a weaker property, relevance, which we state separately because Section 4.5 uses it to derive local decisiveness from a low marginal return to power.

Condition 1 (Relevant and locally decisive cutoff coalition) *There exist a mis-ordered pair $i, j \in I$, with $p_i > p_j$ and $x_i < x_j$, and a scalar $\lambda \in \left(\frac{p_j}{x_j}, \frac{p_i}{x_i}\right)$ such that the cutoff coalition A_λ has the following two properties.*

(i) Relevance. *$P_{A_\lambda} > P_I - (1 - \beta)P_N$.*

(ii) Local decisiveness. *There exists $P_0 \in [0, P_O)$ such that $H_I^{A_\lambda}(P_0) > H_I^A(P_0)$ for every $A \in \mathcal{A}_I \setminus \{A_\lambda\}$.*

By Lemma 2, every mis-ordered pair i, j has a cutoff coalition that contains i but not j and is a best sub-coalition of I. Condition 1 requires that, for at least one mis-ordered pair, such a cutoff coalition is relevant and locally decisive.

Relevance. Part (i) holds if and only if $P_{A_\lambda} + P \geq \beta P_N$ for some $P \in [0, P_O)$, that is, if and only if A_λ can form a winning coalition with a proper outsider coalition. Equivalently, the translated boundary of A_λ is active at some $P \in [0, P_O)$. Relevance excludes cutoff coalitions whose members have high power-to-resource ratios but too little power to form such a winning coalition.

Local decisiveness. Part (ii) requires the translated boundary of A_λ to lie strictly above the translated boundary of every other best sub-coalition at some $P_0 \in [0, P_O)$. At that level of outsider power, an outsider coalition is safe if and only if its resources exceed $H_I^{A_\lambda}(P_0)$. The cutoff coalition A_λ alone therefore determines the external safe area at P_0.

Local decisiveness implies relevance. If the translated boundary of A_λ lies strictly above another translated boundary at P_0, then it is active at P_0, so $P_{A_\lambda} + P_0 \geq \beta P_N$. Since $P_0 < P_O$,

$$P_{A_\lambda} \geq \beta P_N - P_0 > \beta P_N - P_O = P_I - (1-\beta)P_N.$$

We nevertheless state relevance separately, because Proposition 8 derives local decisiveness from relevance and the small-slope condition (16).

Proposition 5 (Strict gains from hierarchy) *Suppose that I is the unique ruling coalition and that Condition 1 holds for a mis-ordered pair $i, j \in I$ and a cutoff coalition A_λ. Consider an admissible sequence of exchanges that begins with an exchange between i and j as in Proposition 4 and then follows the pooling procedure of Corollary 1, and let I' denote the resulting hierarchical allocation. Then I' is strictly more externally resilient than I:*

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

The proof is in Appendix A; we outline its main steps here. The idea is that the first exchange strictly lowers the upper envelope of the translated boundaries at P_0, and, by Proposition 4, the exchanges that follow do not raise it.

Let the pair i, j, the scalar λ, and the cutoff coalition A_λ be as in Condition 1. By Lemma 2, A_λ is a best sub-coalition of I that contains i but not j. By part (ii) of the condition, there exists $P_0 \in [0, P_O)$ at which the translated boundary of A_λ lies strictly above the translated boundary of every other best sub-coalition, so it alone determines the upper envelope at P_0.

Consider the first exchange, between i and j. It leaves the aggregate power and resources of the ruling coalition unchanged, so the indifference curve $H_I(\cdot)$ is also unchanged. Because A_λ contains i but not j, the exchange reduces its power, raises its resources, or both. The translated boundary of A_λ is therefore strictly lower wherever it remains active.

No other translated boundary takes its place. As in the proof of Proposition 4, every sub-coalition after the exchange has weakly less power and weakly more resources than a sub-coalition before it: the one with the same members or, if it contains j' but not i', the one in which i takes the place of j'. Where its translated boundary is active, it therefore lies weakly below the boundary of that earlier sub-coalition. If the earlier sub-coalition is not A_λ, its boundary lies strictly below that of A_λ at P_0. This holds also for sub-coalitions of I that are not best sub-coalitions, because the boundary of each lies strictly below the boundary of some best sub-coalition. The earlier sub-coalition is A_λ in two cases only. The first is A_λ itself after the exchange, whose boundary is strictly lower. The second is the sub-coalition in which j' takes the place of i. Because i and j are mis-ordered, it has strictly less power and strictly more resources than A_λ had before the exchange, so its boundary is also strictly lower. Hence, at P_0, the upper envelope after the exchange lies strictly below the upper envelope before it.

It remains to find a point that is safe after the exchange but not before it. The translated boundary of the full ruling coalition I at P_0 is $H_I(P_I + P_0) - X_I$, which is nonnegative because $H_I(P_I) = X_I$ and H_I is increasing. The upper envelope before the exchange lies strictly above this boundary, so it is positive. We can therefore choose $X > 0$ strictly between the upper envelope after the exchange and the upper envelope before it. The point (P_0, X) is safe after the exchange and unsafe before it. By Proposition 4, every point that is safe before the exchange is also safe after it. Hence the external safe area strictly expands.

Now consider the remaining exchanges in the sequence, which lead to the hierarchical allocation I'. By Proposition 4, each of them weakly expands the external safe area, so the strict gain from the first exchange is preserved. Hence $\mathcal{S}_I^{\text{ext}} \subsetneq \mathcal{S}_{I'}^{\text{ext}}$.[9]

4.4 Heterogeneous preferences

The results so far rely on Assumption 2, under which the members of a coalition rank the coalitions to which they belong in the same way. This subsection relaxes that assumption, considers a single exchange as in Proposition 4, and then extends the result to the sequence of exchanges that leads to a hierarchical allocation (Corollary 1). We keep Assumption 1 and hold each member's plundering-gain function $G_k(\cdot,\cdot)$ fixed during the exchange. We take the incumbent I as given and compare its resilience before and after the exchange. Resilience is now defined with respect to unanimously profitable deviations, i.e., deviations that every member of I who joins them strictly prefers to I.

For each member $k \in I$, let H_k^I denote the indifference curve of k through the incumbent's aggregate power and resources (P_I, X_I). It is defined by $G_k\big(P, H_k^I(P)\big) = G_k(P_I, X_I)$ for every $P \in (0, P_N)$. By Lemma 1, H_k^I is continuously differentiable and strictly increasing.

The following three objects replace the translated boundary, the upper envelope, and the external safe area of Definition 5. For a nonempty sub-coalition $A \subseteq I$, the *unanimity boundary* of A at outsider power $P \in [0, P_O]$ is

$$\underline{H}_I^A(P) := \begin{cases} \min\limits_{k \in A} \left\{ H_k^I(P + P_A) - X_A \right\}, & \text{if } \beta P_N \le P + P_A < P_N, \\ -\infty, & \text{otherwise.} \end{cases} \tag{9}$$

The *external boundary under heterogeneous preferences* is the upper envelope of the unanimity boundaries,

$$\underline{H}_I(P) := \max_{\emptyset \neq A \subseteq I} \underline{H}_I^A(P), \tag{10}$$

and the corresponding external safe area is

$$\mathcal{S}_{I,\text{het}}^{\text{ext}} := \left\{ (P, X) \in \mathcal{D}_O \mid X > \underline{H}_I(P) \right\}. \tag{11}$$

The minimum in Equation (9) expresses unanimity: a deviation is profitable only if every member of I who joins it strictly prefers it to I. Consider a proper winning deviation formed by a sub-coalition $A \subseteq I$ and an outsider coalition with power P and resources X. If $X > \underline{H}_I^A(P)$, at least one member of A strictly prefers I. If $X = \underline{H}_I^A(P)$, every member of A weakly prefers the deviation and at least one is indifferent. The deviation is therefore unanimously profitable if and only if $X < \underline{H}_I^A(P)$.

Under Assumption 2, it was sufficient to consider the best sub-coalitions of I. Here we must consider *every* sub-coalition, because the boundary depends on which members join the deviation, and not only on their aggregate power and resources.

Consider an exchange between i and j as in Proposition 4. Because preferences now differ across members, we keep the names i and j after the exchange and denote the new endowments by

$$p_i' = p_i - \Delta p, \qquad x_i' = x_i + \Delta x, \qquad p_j' = p_j + \Delta p, \qquad x_j' = x_j - \Delta x.$$

[9] As noted in Section 4.1, a strict expansion of the external safe area does not by itself imply a strict increase in the probability that the incumbent survives. That probability strictly increases if and only if $\mu(E_{I'} \setminus E_I) > 0$. A sufficient condition is that $E_{I'} \setminus E_I$ contains a nonempty open subset of $\mathcal{Z}_O$ on which μ has a positive density.

The comparison that matters is between member i before the exchange and member j after it. Define $d_p := p_i - p'_j = p_i - p_j - \Delta p$ and $d_x := x'_j - x_i = x_j - x_i - \Delta x$. Since $\Delta p \leq \frac{p_i - p_j}{2}$ and $\Delta x \leq \frac{x_j - x_i}{2}$, we have $d_p \geq \frac{p_i - p_j}{2} \geq 0$ and $d_x \geq \frac{x_j - x_i}{2} \geq 0$: before the exchange, i has weakly more power and weakly fewer resources than j has after it.

With heterogeneous preferences, the argument in the proof of Proposition 4 breaks down in only one case: sub-coalitions that include j but not i. These are the sub-coalitions that the exchange makes more threatening.

The proof of Proposition 4 compares each such sub-coalition with the sub-coalition before the exchange in which i takes the place of j, which has weakly more power and weakly fewer resources. The comparison now involves two members, i and j, who may have different preferences. Formally, let $A' = C \cup \{j\}$ be the sub-coalition after the exchange, with $C \subseteq I \setminus \{i, j\}$, and let $A = C \cup \{i\}$ be the sub-coalition before it. Then $P_A = P_{A'} + d_p$ and $X_A = X_{A'} - d_x$, where $P_{A'}$ and $X_{A'}$ are the aggregates of A' after the exchange.

We say that the *preference disagreement* between i and j is *bounded* for this exchange if

$$H_j^I(q) \leq H_i^I(q + d_p) + d_x \qquad \text{for every } q \in [\beta P_N, P_N - d_p), \tag{12}$$

where q is the aggregate power of the deviation after the exchange. Under Condition (12), whenever j strictly prefers the deviation after the exchange to I, member i strictly prefers the corresponding coalition before the exchange to I.

Proposition 6 (Exchange under heterogeneous preferences) *Consider an exchange between $i, j \in I$ as in Proposition 4. Suppose that Assumption 1 holds before and after the exchange, and that I is stable against unanimously profitable internal deviations: no proper winning sub-coalition of I is strictly preferred to I by all of its members. Let I' denote the incumbent with its endowments after the exchange. If Condition* (12) *holds, then:*

(i) $\mathcal{S}^{\text{ext}}_{I,\text{het}} \subseteq \mathcal{S}^{\text{ext}}_{I',\text{het}}$*; and*

(ii) I' is stable against unanimously profitable internal deviations.

Suppose in addition that:

(a) i and j are mis-ordered, that is, $p_i > p_j$ and $x_i < x_j$, and the exchange has $\Delta p > 0$ and $\Delta x > 0$;

(b) the inequality in (12) *is strict for every q in its domain; and*

(c) (Local decisiveness) *there exist a sub-coalition $A^* \subseteq I$ that contains i but not j and a point $P_0 \in (0, P_O)$, in the interior of an interval on which the unanimity boundary of A^* is active, such that $\underline{H}_I^{A^*}(P_0) > 0$ and $\underline{H}_I^{A^*}(P_0) > \underline{H}_I^{A}(P_0)$ for every nonempty $A \subseteq I$ with $A \neq A^*$.*

Then $\mathcal{S}^{\text{ext}}_{I,\text{het}} \subsetneq \mathcal{S}^{\text{ext}}_{I',\text{het}}$.

We prove Proposition 6 in Appendix A. The argument has two parts. First, as under Assumption 2, the exchange reduces the power, raises the resources, or both, of any deviation that includes i and excludes j. Such a deviation becomes less attractive to all of its members. Second, a deviation after the exchange that includes j and excludes i must be compared with the coalition before the exchange in which i takes the place of j. That coalition has more power

and fewer resources, but j may be more willing than i to leave I. Condition (12) ensures that the advantage in power and resources outweighs the preference disagreement between i and j.

Condition (12) follows from a simpler condition that involves only two numbers. Let

$$D_{ji} := \sup_{q \in [\beta P_N, P_N - d_p)} \left\{ H_j^I(q) - H_i^I(q) \right\}, \qquad m_i := \inf_{q \in [\beta P_N, P_N)} (H_i^I)'(q).$$

Thus D_{ji} is the largest amount by which the indifference curve of j lies above that of i, and m_i is the smallest slope of the indifference curve of i.

Corollary 2 (Bounded preference disagreement) *Consider an exchange between $i, j \in I$ as in Proposition 4. Suppose that Assumption 1 holds before and after the exchange, that I is stable against unanimously profitable internal deviations, and that D_{ji} and m_i are finite. If*

$$D_{ji} \leq d_x + m_i d_p, \tag{13}$$

then Condition (12) *holds. Hence parts (i) and (ii) of Proposition 6 hold: $\mathcal{S}^{\text{ext}}_{I,\text{het}} \subseteq \mathcal{S}^{\text{ext}}_{I',\text{het}}$, and I' is stable against unanimously profitable internal deviations.*

If, in addition, the inequality in (13) *is strict and Conditions (a) and (c) of Proposition 6 hold, then $\mathcal{S}^{\text{ext}}_{I,\text{het}} \subsetneq \mathcal{S}^{\text{ext}}_{I',\text{het}}$.*

Finally, Inequality (13) *holds whenever $H_i^I(q) \geq H_j^I(q)$ for every $q \in [\beta P_N, P_N - d_p)$, that is, whenever the indifference curve of i lies weakly above that of j.*

Corollary 2 has a simple interpretation. The term D_{ji} bounds how much more willing than i member j can be to join an alternative coalition. The term $d_x + m_i d_p$ measures the advantage of the coalition before the exchange, which contains i, over the corresponding coalition after the exchange, which contains j. The earlier coalition has d_x fewer resources and d_p more power, and each unit of power is worth at least m_i units of resources to i. The exchange weakly expands the external safe area whenever this advantage is at least as large as D_{ji}, the largest amount by which the indifference curve of j lies above that of i. If Inequality (13) fails, the corollary does not apply, although the weaker Condition (12) may still hold.

Proposition 6 concerns a single exchange, but it extends to the pooling sequence of Corollary 1. Suppose that Assumption 1 holds after every exchange in the sequence, that Condition (12) holds for every exchange, and that the incumbent is stable against unanimously profitable internal deviations before the first exchange. Then each exchange preserves this stability and weakly expands the external safe area. The hierarchical allocation is therefore weakly more externally resilient than the initial allocation, as in Corollary 1.

In the remainder of the paper, we again impose Assumption 2, under which the members of a coalition rank coalitions in the same way.

4.5 Gains from hierarchy and the marginal return to power

Proposition 5 shows that an admissible sequence of exchanges leading to a hierarchical allocation strictly increases external resilience if a cutoff coalition A_λ is locally decisive. This subsection relates local decisiveness to the marginal return to power. We show that a low marginal return to power makes relevance easier to satisfy. Given relevance, a low marginal return also implies local decisiveness.

When the marginal return to power is high, a marginal unit of power is equivalent to a large amount of protected resources. The loss of power from excluding any member of the ruling

coalition then exceeds the gain from the resources saved. The most threatening deviation by outsiders therefore includes the full ruling coalition. This threat depends only on the coalition's aggregate power and aggregate resources. Exchanges among its members leave both aggregates unchanged and hence do not affect external resilience.

We measure the marginal return to power by the slope of the indifference curve H_I. Recall that H_I is defined by $G(P, H_I(P)) = G(P_I, X_I)$. Consider a coalition on this curve with power $P \in (\beta P_N, P_N)$ whose power increases by a small amount $\delta > 0$. Its members remain indifferent if its resources, which are protected from plunder, increase by approximately $H_I'(P)\,\delta$, so the slope $H_I'(P)$ measures the increase in protected resources that one additional unit of power offsets. A large value of $H_I'(P)$ corresponds to a high marginal return to power: a small increase in power offsets a large increase in protected resources.

We interpret a low marginal return to power as weak protection of outsiders' property, and a high return as strong protection. When outsiders' property is weakly protected, each unit left outside the coalition yields a large gain, so the marginal return to power is low. When it is strongly protected, a unit yields little unless the coalition is powerful, so the return is high.

Recall that $\overline{H}_I$ denotes the upper envelope of the translated boundaries, defined in (6). Throughout this subsection, we consider outsider coalitions with power $P \in [0, P_O)$.[10]

We call a function G that satisfies Assumptions 1–2 a *plundering technology*. We first compare two plundering technologies, G and $\widetilde{G}$, holding the distribution of power and resources fixed.

Definition 7 (Uniformly higher marginal return to power) *Let H_I and $\widetilde{H}_I$ denote the indifference curves through (P_I, X_I) under the plundering technologies G and $\widetilde{G}$, respectively. The technology $\widetilde{G}$ has a* uniformly higher marginal return to power *than G, denoted $\widetilde{H}_I \succ H_I$, if $\widetilde{H}_I'(P) > H_I'(P)$ for every $P \in (\beta P_N, P_N)$.*

Since both curves pass through (P_I, X_I), we have, for every $P \in [\beta P_N, P_N)$,

$$\widetilde{H}_I(P) - H_I(P) = \int_{P_I}^{P} \left[\widetilde{H}_I'(Q) - H_I'(Q)\right] dQ. \tag{14}$$

If $\widetilde{H}_I \succ H_I$, the integrand in (14) is strictly positive, so the integral is negative for $P < P_I$ and positive for $P > P_I$. Consequently, $\widetilde{H}_I$ lies strictly below H_I on $[\beta P_N, P_I)$ and strictly above it on (P_I, P_N). In the (P, X) plane, a uniformly higher marginal return to power therefore rotates the indifference curve counterclockwise around (P_I, X_I).

4.5.1 The marginal return to power and the ruling coalition

We first show that a uniformly higher marginal return to power leads to a weakly more powerful ruling coalition.

Proposition 7 (Marginal return to power and the power of the ruling coalition) *Fix the distribution of power and resources, and let Z denote the set of potential ruling coalitions. Let G and $\widetilde{G}$ be two plundering technologies. As with G, we write $\widetilde{G}(W) := \widetilde{G}(P_W, X_W)$ for every proper coalition W and set $\widetilde{G}(N) := 0$. Suppose that $I \in \arg\max_{W \in Z} G(W)$ and*

[10] The coalition of all outsiders, O, which is represented by the point (P_O, X_O), is treated separately. At (P_O, X_O), the translated boundary of the ruling coalition I is inactive, because $I \cup O$ is the grand coalition N.

$\widetilde{I} \in \arg\max_{W \in Z} \widetilde{G}(W)$, and let H_I and $\widetilde{H}_I$ denote the indifference curves through (P_I, X_I) under G and $\widetilde{G}$, respectively. If $\widetilde{H}'_I(P) > H'_I(P)$ for every $P \in (\beta P_N, P_I)$, then $P_{\widetilde{I}} \geq P_I$. If, in addition, the endowment profile is generic and $\widetilde{I} \neq I$, then $P_{\widetilde{I}} > P_I$.

The proof is in Appendix A.

To see why, note that both indifference curves pass through (P_I, X_I) and that, on $[\beta P_N, P_I)$, the curve under $\widetilde{G}$ lies strictly below the curve under G. With a higher marginal return to power, a coalition with less power than I thus needs even fewer resources to yield the same plundering gains as I. Under G, no potential ruling coalition yields higher plundering gains than I. Under $\widetilde{G}$, every such coalition with less power than I therefore yields strictly lower plundering gains than I, so the ruling coalition under $\widetilde{G}$ cannot have less power than I.

Proposition 7 implies that a lower marginal return to power makes relevance weakly easier to satisfy. For a proper sub-coalition A of I, the relevance requirement of Condition 1 is

$$P_A > P_I - (1 - \beta) P_N, \tag{15}$$

which holds if and only if A can form a winning coalition with a proper sub-coalition of the outsiders. The right-hand side is increasing in P_I: the more powerful the ruling coalition, the less power remains outside it, and the harder it is for A to form a winning coalition with outsiders. By Proposition 7, a lower marginal return to power leads to a weakly less powerful ruling coalition. The relevance requirement (15) is thus weakly easier to meet for any A that is a proper sub-coalition of the ruling coalition under both technologies.

4.5.2 A low marginal return to power and local decisiveness

We next show that a sufficiently low marginal return to power makes a relevant cutoff coalition locally decisive. Suppose that the best sub-coalitions of I have pairwise distinct aggregate power, and label them in order of decreasing power: $\mathcal{A}_I = \{C_1, \ldots, C_m\}$ with $P_{C_1} > P_{C_2} > \cdots > P_{C_m}$. Then $C_1 = I$, because I has more power than any of its proper sub-coalitions. No best sub-coalition has strictly more power and strictly fewer resources than another, so resources are weakly decreasing in the same order, that is, $X_{C_1} \geq X_{C_2} \geq \cdots \geq X_{C_m}$.

Proposition 8 (A low marginal return to power and local decisiveness) *Suppose that I is the unique ruling coalition and that its best sub-coalitions have pairwise distinct aggregate power, labeled $C_1, \ldots, C_m$ in order of decreasing power. Let $i, j \in I$ be mis-ordered, let $\lambda \in \left(\frac{p_j}{x_j}, \frac{p_i}{x_i}\right)$, and let C_q denote the cutoff coalition A_λ, which is a best sub-coalition of I by Lemma 2. Suppose that C_q is relevant, that is, $P_{C_q} > P_I - (1 - \beta) P_N$. If there exists $\bar{h} > 0$ such that $H'_I(P) \leq \bar{h}$ for every $P \in (\beta P_N, P_N)$ and*

$$\bar{h} < \min \left\{ \min_{r<q} \frac{X_{C_r} - X_{C_q}}{P_{C_r} - P_{C_q}},\ \min_{s>q} \frac{X_I - X_{C_s}}{P_I - P_{C_s}} \right\}, \tag{16}$$

where a minimum over an empty set equals $+\infty$, then C_q is locally decisive: there exists $P_0 \in [0, P_O)$ at which the translated boundary of C_q lies strictly above the translated boundary of every other best sub-coalition.

The proof is in Appendix A. The bound in (16) plays two roles in the argument. First, it implies that $P_{C_s} < \beta P_N$ for every $s \geq q$: neither C_q nor any best sub-coalition with less power

is winning on its own. If one of them were winning, it would yield higher plundering gains than I, because under the bound its fewer resources more than offset its lower power. This would contradict the assumption that I is the unique ruling coalition. Because C_q is relevant, it can form a winning coalition with outsiders of total power $P < P_O$. Since the best sub-coalitions have distinct power, we can choose $P_0 \in [0, P_O)$ such that $P_{C_q} + P_0 \geq \beta P_N$ and $P_{C_s} + P_0 < \beta P_N$ for every $s > q$. Hence, at P_0, the translated boundary of C_q is active, and the translated boundary of every best sub-coalition with less power is inactive.

Second, the bound implies that, at P_0, the translated boundary of every best sub-coalition with more power than C_q lies strictly below that of C_q. A unit of power offsets at most $\bar{h}$ units of resources, whereas, by (16), such a sub-coalition holds more than $\bar{h}$ additional units of resources for each additional unit of power. As a result, its additional power does not offset its additional resources. Hence, at P_0, the translated boundary of C_q is active and lies strictly above the translated boundary of every other best sub-coalition, so C_q is locally decisive.

Together, Propositions 7 and 8 show how a low marginal return to power helps to satisfy the two requirements of Condition 1. First, a lower marginal return to power leads to a weakly less powerful ruling coalition, which leaves more power outside it. A given sub-coalition then forms a winning coalition with outsiders more easily, so relevance is easier to satisfy. Second, under Condition (16), a relevant cutoff coalition is locally decisive. In that case, Proposition 5 implies that an admissible sequence of exchanges that begins with an exchange between i and j and then follows the pooling procedure of Corollary 1 strictly increases external resilience.

4.5.3 A high marginal return to power and no gain from hierarchy

When the marginal return to power is sufficiently high, admissible exchanges leave external resilience unchanged. The threshold is

$$\Psi_I := \max_{A \in \mathcal{A}_I \setminus \{I\}} \frac{X_I - X_A}{P_I - P_A}, \tag{17}$$

with $\Psi_I := 0$ if I has no proper best sub-coalition. For a proper best sub-coalition A, the ratio $\frac{X_I - X_A}{P_I - P_A}$ is the amount of resources saved per unit of power lost when A replaces I in a deviation, and Ψ_I is the largest such ratio. We say that the marginal return to power is *high* if there exists $\psi > \Psi_I$ such that $H_I'(P) \geq \psi$ for every $P \in (\beta P_N, P_N)$.

Proposition 9 (A high marginal return to power and no gain from exchanges) *Suppose that I is the unique ruling coalition and that the marginal return to power is high, that is, there exists $\psi > \Psi_I$ such that $H_I'(P) \geq \psi$ for every $P \in (\beta P_N, P_N)$. Then:*

(i) the full ruling coalition determines the upper envelope, that is,

$$\overline{H}_I(P) = H_I(P + P_I) - X_I \qquad \text{for every } P \in [0, P_O);$$

(ii) an admissible exchange as in Proposition 4 leaves external resilience unchanged: if I' denotes the coalition after the exchange, then $\mathcal{S}^{\text{ext}}_{I'} = \mathcal{S}^{\text{ext}}_I$.

The proof is in Appendix A. To see the idea behind part (i), recall that Ψ_I is the largest amount of resources saved per unit of power lost when a proper best sub-coalition replaces I in a deviation. When the marginal return to power is high, the power lost is worth more than

the resources saved. Hence a proper sub-coalition of I is never a more threatening partner for outsiders than I itself. The full ruling coalition therefore determines the upper envelope on $[0, P_O)$. The binding deviation thus enlarges the ruling coalition, i.e., it admits outsiders and keeps all of its members.

For part (ii), note that an exchange may change which sub-coalitions are best sub-coalitions, but it leaves (P_I, X_I) unchanged. As in the proof of Proposition 4, every best sub-coalition after the exchange has weakly less power and weakly more resources than some best sub-coalition before it. As a result, the threshold does not rise, i.e., $\Psi_{I'} \leq \Psi_I$, and the translated boundary of the full ruling coalition still determines the upper envelope after the exchange. This boundary depends only on (P_I, X_I), so the external safe area is unchanged for $P \in [0, P_O)$. The point (P_O, X_O) is safe under both allocations by Proposition 4, and hence the two external safe areas coincide.

Proposition 9 thus shows that, when the marginal return to power is high, the weak inclusion in Proposition 4 holds with equality.

4.6 A trade-off between internal and external resilience

Our final result states that, when the marginal return to power is high, raising it further weakly increases internal resilience and strictly reduces external resilience.

We compare two plundering technologies, G and $\widetilde{G}$, for a fixed coalition I that is the ruling coalition under G. Let H_I and $\widetilde{H}_I$ denote the indifference curves through (P_I, X_I) under G and $\widetilde{G}$, respectively. We apply Definitions 4 and 5 to the same set of members I under both technologies, whether or not I is the ruling coalition under $\widetilde{G}$. Let $\mathcal{S}_I^{\text{int}}(H_I)$ and $\mathcal{S}_I^{\text{ext}}(H_I)$ denote the internal and external safe areas under G, and let $\mathcal{S}_I^{\text{int}}(\widetilde{H}_I)$ and $\mathcal{S}_I^{\text{ext}}(\widetilde{H}_I)$ denote those under $\widetilde{G}$.

Every proper sub-coalition of I has less power than I, so only the part of the internal safe area with $P < P_I$ matters for internal deviations. We denote this part by

$$\mathcal{S}_I^{\text{int},-}(H_I) := \mathcal{S}_I^{\text{int}}(H_I) \cap \{(P, X) : P < P_I\},$$

and define $\mathcal{S}_I^{\text{int},-}(\widetilde{H}_I)$ analogously.

Proposition 10 (A trade-off between internal and external resilience) *Let I be a proper coalition that is the ruling coalition under G, and let H_I and $\widetilde{H}_I$ denote the indifference curves through (P_I, X_I) under G and $\widetilde{G}$, respectively. Suppose that:*

(a) the marginal return to power is high under G, that is, there exists $\psi > \Psi_I$ such that $H_I'(P) \geq \psi$ for every $P \in (\beta P_N, P_N)$;

(b) $\widetilde{G}$ has a uniformly higher marginal return to power than G, that is, $\widetilde{H}_I'(P) > H_I'(P)$ for every $P \in (\beta P_N, P_N)$;

(c) the point (P_O, X_O) is safe under both technologies.

Then internal resilience is weakly higher under $\widetilde{G}$,

$$\mathcal{S}_I^{\text{int},-}(H_I) \subseteq \mathcal{S}_I^{\text{int},-}(\widetilde{H}_I), \tag{18}$$

and external resilience is weakly lower under $\widetilde{G}$*,*

$$\mathcal{S}_I^{\text{ext}}(\widetilde{H}_I) \subseteq \mathcal{S}_I^{\text{ext}}(H_I). \tag{19}$$

The inclusion in (18) *is strict if* $\beta P_N < P_I$*, and the inclusion in* (19) *is always strict.*

Appendix A contains the proof. A proper winning sub-coalition of I has power $P \in [\beta P_N, P_I)$. On this interval, a uniformly higher marginal return to power lowers the indifference curve through (P_I, X_I). So a sub-coalition must have fewer resources for its members to prefer secession to I. Hence the internal safe area restricted to $P < P_I$ weakly expands.

On the other hand, the external safe area shrinks. By Conditions (a) and (b) and Proposition 9, the full ruling coalition determines the upper envelope under both technologies. The most threatening deviation by an outsider coalition with power $P \in (0, P_O)$ is therefore to join the full ruling coalition I, which gives a coalition with power $P + P_I \in (P_I, P_N)$. On this interval, a uniformly higher marginal return to power raises the indifference curve through (P_I, X_I), and with it the translated boundary of the full ruling coalition, $H_I(P + P_I) - X_I$. An outsider coalition must then have more resources than under G to be safe, so the external safe area shrinks for $P \in (0, P_O)$. The coalition of all outsiders is safe under both technologies by Condition (c).

The comparison of external safe areas in (19) relies on Condition (a), which, together with Condition (b), ensures that the same insider partner, the full ruling coalition, determines the upper envelope under both technologies. Without this condition, a proper sub-coalition of I may determine the upper envelope. A higher marginal return to power raises the translated boundaries of deviations with total power above P_I and lowers those of deviations with total power below P_I. It may therefore change which translated boundary determines the upper envelope, and its effect on external resilience is ambiguous in general.

4.7 Coalitional resilience: a summary

Table 1 summarizes the results of this section. An admissible sequence of exchanges that leads to a hierarchical allocation weakly increases external resilience, regardless of the marginal return to power (Corollary 1). The increase is strict when the marginal return to power is low and a cutoff coalition is relevant (Propositions 5 and 8), and there is no increase when the marginal return to power is high (Proposition 9). When the marginal return to power is high, raising it further weakly increases internal resilience and strictly reduces external resilience (Proposition 10).

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 coalitions with relatively high power and few resources can be particularly threatening. This offers a resource-based perspective on 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 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. At

	Low marginal return to power	**High marginal return to power**
Effect of hierarchy on external resilience	An admissible sequence of exchanges leading to a hierarchical allocation weakly increases external resilience (Corollary 1).	
Is the gain strict?	Yes, if a cutoff coalition is relevant, for a sequence constructed as in Corollary 1 (Propositions 5 and 8).	No. Admissible exchanges leave external resilience unchanged (Proposition 9).
Trade-off between internal and external resilience	No general result.	A uniformly higher marginal return to power weakly increases internal resilience and strictly reduces external resilience (Proposition 10).

Table 1: Summary of the results on resilience for a low and a high marginal return to power. The strict gain holds for the sequence of exchanges described in Proposition 5. The two cases are not exhaustive.

a given level of power, the relative poverty of nomadic tribes compared to urban civilizations gives them a high power-to-resource ratio. In our model, this can make such groups attractive partners in a coalition to plunder cities, a repeated pattern in the pre-modern world. Similar logic might also help understand communist revolutions (Morishima, 1974; Roemer, 1980, 1981; Brewer, 2002), where relatively poor groups can be attractive coalition partners if they retain substantial power. The same logic applies inside such a coalition once it rules. Members who remain strong but poor can be the most attractive partners for the next group of outsiders. Aligning resources with power, so that no member has more power and fewer resources than another, removes this threat. This may help explain why successful plundering coalitions come to match wealth to rank. Our results do not explain how unequal ranks emerge from full equality: an allocation in which all members are identical is already hierarchical in our sense, and our exchanges start from a misalignment between power and resources.

Although the model is static, the resilience object also provides a natural one-period building block for dynamic environments. It isolates two forces that shape the evolution of ruling coalitions over time: how the distribution of power and resources affects internal cohesion, and how it affects exposure to outsider challenges. The static analysis therefore offers a benchmark for dynamic questions about coalition survival and institutional change.

A natural next step is to endogenize the objects that are exogenous here. One extension would allow players to invest in power before coalition formation. This would clarify how initial resources shape incentives to accumulate coercive capacity and, ultimately, which coalition rules (Acemoglu et al., 2008). A broader extension would endogenize the plundering technology itself. In a dynamic version of the model, ruling coalitions could invest in institutions that change

how power translates into extraction. This would connect the analysis to work on coalition formation, institutional choice, and credible commitment under political conflict (Hafer, 2006; Besley and Persson, 2009; Gehlbach and Keefer, 2011; Diermeier et al., 2017; Cao and Lagunoff, 2023). It would also allow the framework to speak to the persistence of early institutional choices and their long-run political and economic consequences (Michalopoulos and Papaioannou, 2013; Lowes et al., 2017; Chambru et al., 2024).

References

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

Daron Acemoglu, Davide Ticchi, and Andrea Vindigni. A theory of military dictatorships. *American Economic Journal: Macroeconomics*, 2(1):1–42, 2010.

Daron Acemoglu, Georgy Egorov, and Konstantin Sonin. Dynamics and stability of constitutions, coalitions, and clubs. *American Economic Review*, 102(4):1446–1476, 2012.

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

M. Athar Ali. *The Mughal Nobility under Aurangzeb*. Asia Publishing House, Bombay, 1966.

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

Timothy Besley and James A. Robinson. Quis custodiet ipsos custodes? civilian control over the military. *Journal of the European Economic Association*, 8(2–3):655–663, 2010. doi: 10.1111/j.1542-4774.2010.tb00535.x.

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

Carles Boix and Milan W Svolik. The foundations of limited authoritarian government: Institutions, commitment, and power-sharing in dictatorships. *the Journal of Politics*, 75(2): 300–316, 2013.

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

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

Dan Cao and Roger Lagunoff. The dynamics of property rights in modern autocracies. *American Economic Journal: Microeconomics*, 15(3):305–353, 2023.

Cédric Chambru, Emeric Henry, and Benjamin Marx. The dynamic consequences of state building: Evidence from the french revolution. *American Economic Review*, 114(11):3578–3622, 2024.

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

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

Georgy Egorov and Konstantin Sonin. The political economics of non-democracy. *Journal of Economic Literature*, 62(2):594–636, 2024.

Joan Esteban and Debraj Ray. On the salience of ethnic conflict. *American Economic Review*, 98(5):2185–2202, 2008.

Patrick Francois, Ilia Rainer, and Francesco Trebbi. How is power shared in africa? *Econometrica*, 83(2):465–503, 2015.

Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of political economy*, 108(5):874–904, 2000.

Luis Garicano and Esteban Rossi-Hansberg. Knowledge-based hierarchies: Using organizations to understand the economy. *Annual Review of Economics*, 7:1–30, 2015. doi: 10.1146/annurev-economics-080614-115748.

Scott Gehlbach and Philip Keefer. Investment without democracy: Ruling-party institutionalization and credible commitment in autocracies. *Journal of Comparative Economics*, 39(2): 123–139, 2011.

Scott Gehlbach, Konstantin Sonin, and Milan W Svolik. Formal models of nondemocratic politics. *Annual Review of Political Science*, 19:565–584, 2016.

Herschel I. Grossman and Minseong Kim. Swords or plowshares? A theory of the security of claims to property. *Journal of Political Economy*, 103(6):1275–1288, 1995.

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

Soeren J. Henn, Christian Mastaki Mugaruka, Miguel Ortiz, Raúl Sánchez de la Sierra, and David Qihang Wu. Monopoly of taxation without a monopoly of violence: The weak state's trade-offs from taxation. *The Review of Economic Studies*, 92(2):1126–1156, 2025. doi: 10.1093/restud/rdae036.

Jack Hirshleifer. Anarchy and its breakdown. *Journal of Political Economy*, 103(1):26–52, 1995.

Halil İnalcık. *The Ottoman Empire: The Classical Age, 1300–1600*. Weidenfeld and Nicolson, London, 1973.

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

James S. Jordan. Pillage and property. *Journal of Economic Theory*, 131(1):26–44, 2006. doi: 10.1016/j.jet.2005.05.008.

Ibn Khaldun. The muqaddimah, an introduction to history, tr. by franz rosenthal, 3 vols, 1967.

Mert Kimya. Power, status, and the stability of hierarchies. Working paper, University of Sydney, May 4, 2026.

Kai A. Konrad and Stergios Skaperdas. The market for protection and the origin of the state. *Economic Theory*, 50(2):417–443, 2012.

Timur Kuran. The discontents of islamic economic morality. *The American Economic Review*, 86(2):438–442, 1996.

Sara Lowes, Nathan Nunn, James A Robinson, and Jonathan L Weigel. The evolution of culture and institutions: Evidence from the kuba kingdom. *Econometrica*, 85(4):1065–1091, 2017.

Simon MacKenzie, Manfred Kerber, and Colin Rowat. Pillage games with multiple stable sets. *International Journal of Game Theory*, 44(4):993–1013, 2015.

Joram Mayshar, Omer Moav, and Luigi Pascali. The origin of the state: Land productivity or appropriability? *Journal of Political Economy*, 130(4):1091–1144, 2022.

Anne Meng. *Constraining dictatorship*. Cambridge University Press Cambridge, 2020.

Stelios Michalopoulos and Elias Papaioannou. Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152, 2013.

Dilip Mookherjee. Incentives in hierarchies. *The Handbook of Organizational Economics*, pages 764–798, 2013. edited by Robert Gibbons and John Roberts.

W. H. Moreland. Rank (mansab) in the mogul state service. *Journal of the Royal Asiatic Society of Great Britain and Ireland*, pages 641–665, 1936.

Michio Morishima. Marx in the light of modern economic theory. *Econometrica: Journal of the Econometric Society*, pages 611–632, 1974.

Douglass C. North, John Joseph Wallis, and Barry R. Weingast. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge University Press, 2009.

Mancur Olson. Dictatorship, democracy, and development. *American political science review*, 87(3):567–576, 1993.

Jack Paine. The dictator's power-sharing dilemma: Countering dual outsider threats. *American Journal of Political Science*, 65(2):510–527, 2021.

Michele Piccione and Ronny Razin. Coalition formation under power relations. *Theoretical Economics*, 4(1):1–15, 2009.

Michele Piccione and Ariel Rubinstein. Equilibrium in the jungle. *The Economic Journal*, 117 (522):883–896, 2007. doi: 10.1111/j.1468-0297.2007.02072.x.

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

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

Debraj Ray and Rajiv Vohra. Coalition formation. In H. Peyton Young and Shmuel Zamir, editors, *Handbook of Game Theory with Economic Applications*, volume 4, chapter 5, pages 239–326. North-Holland, 2015. doi: 10.1016/B978-0-444-53766-9.00005-7.

William H. Riker. *The Theory of Political Coalitions*. Yale University Press, New Haven, 1962.

John E Roemer. A general equilibrium approach to marxian economics. *Econometrica: Journal of the Econometric Society*, pages 505–530, 1980.

John E Roemer. *Analytical foundations of Marxian economic theory*. Cambridge University Press, 1981.

Philip Roessler. The enemy within: Personal rule, coups, and civil war in Africa. *World Politics*, 63(2):300–346, 2011.

Raúl Sánchez de la Sierra. On the origins of the state: Stationary bandits and taxation in Eastern Congo. *Journal of Political Economy*, 128(1):32–74, 2020. doi: 10.1086/703989.

Stergios Skaperdas. Cooperation, conflict, and power in the absence of property rights. *American Economic Review*, 82(4):720–739, 1992.

Milan W Svolik. Power sharing and leadership dynamics in authoritarian regimes. *American Journal of Political Science*, 53(2):477–494, 2009.

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

Charles Tilly. War making and state making as organized crime. In Peter B. Evans, Dietrich Rueschemeyer, and Theda Skocpol, editors, *Bringing the State Back In*, pages 169–191. Cambridge University Press, Cambridge, 1985.

Alexis de Tocqueville. *L'Ancien Régime et la Révolution*. Michel Lévy Frères, Paris, 1856.

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

Ronald Wintrobe. *The political economy of dictatorship*. Cambridge University Press, 2000.

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

Appendix

A Proofs

Proof of Lemma 1. Fix $i \in N$, a plundering gain $u > 0$, and $P \in (0, P_N)$. By Assumption 1, $G_i(P, \cdot)$ is continuous and strictly decreasing, with limits $+\infty$ as $X \to 0^+$ and 0 as $X \to \infty$. Hence there exists a unique $X = H_i(P; u) > 0$ such that $G_i(P, H_i(P; u)) = u$. Thus the level set is the graph of a function on all $P \in (0, P_N)$.

If $P' > P$, then $G_i(P', H_i(P; u)) > G_i(P, H_i(P; u)) = u$. Since $G_i(P', \cdot)$ is strictly decreasing, holding the gain fixed at u requires $H_i(P'; u) > H_i(P; u)$. Hence the indifference curve is strictly increasing.

Finally, G_i is continuously differentiable and $\partial G_i/\partial X < 0$. The implicit-function theorem therefore gives $H_i \in C^1$ and

$$H_i'(P; u) = -\frac{\partial G_i/\partial P}{\partial G_i/\partial X}\big(P, H_i(P; u)\big) > 0.$$

■

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 nonempty, $\arg\max_{W \in \mathcal{W}} G(W)$ is well-defined and nonempty, so $\phi(I) \neq \emptyset$, establishing the first part of Axiom 1. Moreover, by construction, $\phi(I) \subseteq \mathcal{W}$, so Axiom 2 holds.

For $I' \in \phi(I)$ and $I'' \in \mathcal{W}$, we have $I'' \notin \phi(I)$ if and only if $G(I'') < G(I')$, because I' maximizes G over $\mathcal{W}$. This is exactly Axiom 3. Finally, by convention, $G(N) = 0$, 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 $I' \in \phi'(I)$, which exists by Axiom 1. Any other selected coalition has the same value of G: applying Axiom 3 in both directions rules out either strict inequality. Every unselected winning coalition has strictly lower G by the same axiom. Thus the selected coalitions are exactly the maximizers of G over $\mathcal{W}$, so $\phi' = \phi$.

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, $C = \phi(I_0)$. Fix any $I \in C$ and let $K := I \cap I_0$. Since both I and I_0 are winning and $\beta > \frac{1}{2}$, we have $K \neq \emptyset$.

We first record a payoff comparison. Every member of I weakly prefers I to any possible ruling coalition. If $i \in I \cap W$ and $W \subsetneq N$, then Assumption 2 and the fact that I maximizes G imply $U_i(W) = x_i + g(i)G(W) \leq x_i + g(i)G(I) = U_i(I)$. If $W = N$, then $U_i(N) = x_i < U_i(I)$; if $i \notin W$, then $U_i(W) < x_i < U_i(I)$ by Assumption 1.

We construct an SPE by backward induction. At any agenda-setting history at which no member of K remains eligible to propose, use a pure-strategy SPE of the remaining finite game. At any voting history, votes are chosen by backward induction given the continuation following rejection. Thus a voter accepts when doing so is optimal and rejects when rejection is optimal; either action may be selected when she is indifferent.

Now consider an agenda-setting history at which at least one member of K remains. If the current agenda setter belongs to K, let her propose I. Every member of I weakly prefers I to

the continuation following rejection, so we may select acceptance at every voting node and I forms. Proposing I is optimal for the agenda setter by the payoff comparison above.

Suppose instead that the current agenda setter does not belong to K and that some member of K remains for a later round. By the induction hypothesis, rejection of the current proposal leads to I. Consider any winning proposal $W \neq I$. Since W and I are both winning and $\beta > \frac{1}{2}$, they intersect. Every member of $W \cap I$ weakly prefers the continuation outcome I to W. We can therefore select an equilibrium of the voting subgame in which one such member rejects W whenever necessary, with indifference resolved in favor of rejection. Hence no winning proposal different from I is accepted. A proposal of I is accepted, while a non-winning or rejected proposal leads to the same continuation outcome I. The continuation outcome is therefore I.

This induction constructs a pure-strategy SPE yielding I. Notice that the argument does not prescribe rejection after every off-path proposal. In particular, if a member of K is the last such agenda setter and proposes some $W \neq I$, votes are determined by backward induction given the continuation after rejection. This does not affect the equilibrium path because that agenda setter weakly prefers proposing I. ■

Proof of Proposition 2(2). Let $K := I \cap I_0$. Since I and I_0 are winning and $\beta > \frac{1}{2}$, we have $K \neq \emptyset$. Since $\phi(I_0) = \{I\}$, Proposition 1 implies that I is the unique maximizer of G over $\mathcal{W}$. Under Assumptions 1–2, every member of I therefore strictly prefers I to any other possible ruling coalition. If an alternative coalition contains that member, uniqueness of the maximizer of G gives the strict comparison. If it excludes that member, Assumption 1(2) gives a payoff below her initial resources, whereas membership in I gives a strictly positive gain.

First consider a voting subgame after a proposal of I. Backward induction through the sequential votes shows that every SPE yields I: a voter accepts whenever rejection would lead to a different outcome, and can reject only when the continuation already yields I. Thus proposing I guarantees outcome I, although acceptance by every voter need not occur.

We prove uniqueness of the equilibrium *outcome* by backward induction on the number of remaining agenda-setting rounds, restricting attention to histories at which at least one member of K remains available. If only one round remains, its agenda setter must belong to K. By proposing I, she can guarantee her strictly preferred ruling-coalition outcome, as the voting argument above shows. Hence every SPE continuation outcome from such a history is I.

Now suppose the claim holds whenever fewer rounds remain, and consider an earlier history at which at least one member of K remains available. If the current agenda setter belongs to K, she can again guarantee outcome I by proposing it. Since she strictly prefers I to every other possible ruling-coalition outcome, every sequentially rational action of hers must induce I; equilibrium strategies may differ only through actions whose continuation outcome is still I.

If instead the current agenda setter does not belong to K, rejection of the current proposal leaves every remaining member of K available. The induction hypothesis therefore implies continuation outcome I. Consider any winning proposal $J \neq I$. Because $\beta > \frac{1}{2}$, two winning coalitions cannot be disjoint, so $J \cap I \neq \emptyset$. A voter in $J \cap I$ strictly prefers the continuation outcome I to J and rejects whenever her acceptance would allow J to form. Earlier rejection may already deliver I. Non-winning proposals fail by the rules of the game. Thus the current agenda setter cannot induce a ruling-coalition outcome different from I, and the continuation outcome is again I.

Backward induction implies that every SPE produces I as the ruling coalition, although equilibrium strategies need not be unique. ■

Proof of Proposition 3. We prove both directions.

(⇒) 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 $W \in \mathcal{W}\backslash\{I\}$.

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).

(⇐) 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 $W \in \mathcal{W} \setminus \{I\}$.

If $I = N$, a proper winning coalition exists by assumption. Repeatedly replacing it by a coalition with strictly more power and strictly fewer resources, when possible, gives a proper winning member of $\mathcal{A}_N$ with strictly positive G. This contradicts Condition (i). Thus I is proper and $G(I) > G(N) = 0$. Fix any proper $W \in \mathcal{W} \setminus \{I\}$ and write $B := W \cap I$ and $C := W \setminus I$. Since W and I are winning and $\beta > \frac{1}{2}$, $B \neq \emptyset$. 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 $\widetilde{A}^{\text{ins}} \in \mathcal{W}$. Moreover, $X_{\widetilde{A}^{\text{ins}}} \leq X_W < X_I$ because $W \subsetneq I$ and all resources are positive. Hence $\widetilde{A}^{\text{ins}} \neq I$.

Condition (i) therefore gives $G(I) > G(\widetilde{A}^{\text{ins}})$. Since G is increasing in power and decreasing in resources throughout its domain, $P_{\widetilde{A}^{\text{ins}}} \geq P_W$ and $X_{\widetilde{A}^{\text{ins}}} \leq X_W$ imply $G(\widetilde{A}^{\text{ins}}) \geq G(W)$. 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 $\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}} \in \mathcal{W}$. Moreover, $X_{\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}} \leq X_W < X_N$, so the union is proper. By Condition (ii), $G(I) > G(\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}})$.

By the same monotonicity of G, the aggregate inequalities imply $G(\widetilde{A}^{\text{ins}} \cup \widetilde{A}^{\text{ext}}) \geq G(W)$. 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 $I' := (I \setminus \{i, j\}) \cup \{i', j'\}$ denote 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 $0 \leq \Delta p \leq \frac{p_i - p_j}{2}$, $0 \leq \Delta x \leq \frac{x_j - x_i}{2}$, and $(\Delta p, \Delta x) \neq (0, 0)$. Then $p_{i'} \geq p_{j'}$ and $x_{i'} \leq x_{j'}$. The exchange preserves the aggregate power and resources of the ruling coalition, so $P_{I'} = P_I$ and $X_{I'} = X_I$. Hence the indifference curve through the ruling coalition is unchanged; denote it by H_I.

Use H_I^A and $H_{I'}^{A'}$ for the translated boundaries of nonempty sub-coalitions $A \subseteq I$ and $A' \subseteq I'$, given by the formula in Definition 5 before and after the exchange, respectively.

We begin with a comparison claim for best sub-coalitions before and after the exchange.

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

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

If $i', j' \in A'$, let $A^- := (A' \setminus \{i', j'\}) \cup \{i, j\}$. Then $P_{A^-} = P_{A'}$ and $X_{A^-} = X_{A'}$.

If $i', j' \notin A'$, let $A^- := A'$. Again, $P_{A^-} = P_{A'}$ and $X_{A^-} = X_{A'}$.

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 $P_{A^-} = P_{A'} + \Delta p \geq P_{A'}$ and $X_{A^-} = X_{A'} - \Delta x \leq X_{A'}$.

If $j' \in A'$ and $i' \notin A'$, let $A^- := (A' \setminus \{j'\}) \cup \{i\}$. Then $P_{A^-} - P_{A'} = p_i - p_{j'} = p_i - p_j - \Delta p \geq 0$ and $X_{A^-} - X_{A'} = x_i - x_{j'} = x_i - x_j + \Delta x \leq 0$ by the bounds on $(\Delta p, \Delta x)$. Hence $P_{A^-} \geq P_{A'}$ and $X_{A^-} \leq X_{A'}$.

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

If $A^- \in \mathcal{A}_I$, set $A := A^-$. Otherwise, since $A^- \notin \mathcal{A}_I$, there exists $B \subseteq I$ such that $P_B > P_{A^-}$ and $X_B < X_{A^-}$. If $B \in \mathcal{A}_I$, set $A := B$. If not, iterate the same procedure. This generates a sequence of sub-coalitions of I with strictly increasing power and strictly decreasing resources. Since I is finite, the procedure 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 expands the external safe area.

Since the exchange only redistributes power and resources among members of the ruling coalition, $P_{I'} = P_I$ and (P_O, X_O) is unchanged. Hence $\mathcal{D}_O$ is unchanged. Fix an arbitrary point $(P, X) \in \mathcal{S}_I^{\text{ext}}$ and any $A' \in \mathcal{A}_{I'}$. By the claim, there exists $A \in \mathcal{A}_I$ such that $P_A \geq P_{A'}$ and $X_A \leq X_{A'}$.

If $P + P_{A'} < \beta P_N$ or $P + P_{A'} = P_N$, then $H_{I'}^{A'}(P) = -\infty$ and the constraint is automatic. Otherwise, $\beta P_N \leq P + P_{A'} < P_N$ and $P + P_A \geq P + P_{A'}$. If $P < P_O$, then $P + P_A < P_N$. If $P = P_O$, finiteness of $H_{I'}^{A'}(P)$ implies $A' \subsetneq I'$, and $X_A \leq X_{A'} < X_I$ implies $A \subsetneq I$, so again $P + P_A < P_N$. Since H_I is increasing, $H_{I'}^{A'}(P) = H_I(P + P_{A'}) - X_{A'} \leq H_I(P + P_A) - X_A = H_I^A(P)$. Because $(P, X) \in \mathcal{S}_I^{\text{ext}}$, we have $X > H_I^A(P)$, and hence $X > H_{I'}^{A'}(P)$. As A' was arbitrary, $(P, X) \in \mathcal{S}_{I'}^{\text{ext}}$. Therefore $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{I'}^{\text{ext}}$.

It remains to verify that internal stability is preserved after the exchange. Let $A' \in (\mathcal{A}_{I'} \setminus \{I'\}) \cap \mathcal{W}$ be any proper winning best 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 $A \in \mathcal{A}_I \cap \mathcal{W}$. Moreover, $A \neq I$: because $A' \subsetneq I'$ and all resources are positive, $X_{A'} < X_{I'} = X_I$, whereas $X_A \leq X_{A'}$.

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 Condition (i) of Proposition 3 continues to hold after the exchange.

Therefore, $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{I'}^{\text{ext}}$ and internal stability is preserved. Since outsider endowments are unchanged and $\mathcal{S}_I^{\text{ext}} \subseteq \mathcal{S}_{I'}^{\text{ext}}$, Condition (ii) of Proposition 3, which holds for I, also holds for I'. Proposition 3, which does not require Assumption 3, therefore implies that $\phi(I') = \{I'\}$ after the exchange. Proposition 2 then implies that I' is the unique ruling coalition in every SPE. Hence the exchange weakly increases the external resilience of the ruling coalition. ■

The number of ranks. Order the members of I as at the start of the pooling procedure, from most to least powerful, with ties broken in favor of more resources, and let $X^{(k)}$ denote the total resources of the first k members. Suppose that I contains a mis-ordered pair. Then the procedure ends with a single rank, in which every member has power $\frac{P_I}{|I|}$ and resources $\frac{X_I}{|I|}$, if and only if

$$\frac{X^{(k)}}{k} < \frac{X_I}{|I|} \qquad \text{for every } k \in \{1, \ldots, |I| - 1\}. \tag{20}$$

That is, full equality is reached if and only if every group of the most powerful members holds strictly less than its per-capita share of the coalition's resources. If Condition (20) fails for some k, the hierarchical allocation has at least two ranks. This is the case, for example, whenever the most powerful member holds at least the average resources of a member ($k = 1$), or the least powerful member holds at most the average ($k = |I| - 1$).

Proof of the claim on the number of ranks (Condition (20)). Order the members of I as at the start of the pooling procedure. Every block is a set of consecutive members in this order, and pooling replaces two adjacent blocks by their union. For a set B of consecutive members, let $\bar{x}_B$ denote its resources per member. An *initial segment* of a block consists of its first members in the order, and it is *proper* if it omits at least one member of the block.

We first show that every block B formed during the procedure has the following property: $\bar{x}_S < \bar{x}_B$ for every proper initial segment S of B. The property holds trivially for a block with one member. Suppose that it holds for two adjacent blocks A and B that are pooled, with A first. Because the two blocks are mis-ordered, $\bar{x}_A < \bar{x}_B$, so $\bar{x}_A < \bar{x}_{A\cup B} < \bar{x}_B$. Let S be a proper initial segment of $A \cup B$. If $S \subseteq A$, then $\bar{x}_S \leq \bar{x}_A < \bar{x}_{A\cup B}$. Otherwise, $S = A \cup T$ for a proper initial segment T of B. Then $\bar{x}_T < \bar{x}_B$, so the remaining members of B satisfy $\bar{x}_{B\setminus T} > \bar{x}_B > \bar{x}_{A\cup B}$. Since $\bar{x}_{A\cup B}$ is a weighted average of $\bar{x}_S$ and $\bar{x}_{B\setminus T}$, it follows that $\bar{x}_S < \bar{x}_{A\cup B}$.

Suppose that the procedure ends with a single rank. Because I contains a mis-ordered pair, at least one pooling step occurs. When two blocks are pooled, the power per member of the pooled block lies strictly between theirs. It is therefore strictly below that of the preceding block and strictly above that of the following block, and later pooling steps preserve both strict inequalities. Hence a pooled block never has the same endowment as an adjacent block, and all members can have the same endowment only if the procedure ends with a single block, which is I. Applying the property above to this block gives Condition (20).

Conversely, suppose that Condition (20) holds and that the procedure ends with at least two blocks. Let B_1 denote the first block, with $k < |I|$ members. When the procedure ends, resources per member weakly decrease along the ordering. Hence $\bar{x}_{B_1}$ is at least the resources per member of every other block, and therefore at least $\frac{X_I}{|I|}$, which is a weighted average of the resources per member of all blocks. That is, $\frac{X^{(k)}}{k} \geq \frac{X_I}{|I|}$, which contradicts Condition (20). The procedure therefore ends with a single block, in which all members have the same endowment. ■

Optimal design of the internal allocation. Consider a designer who chooses the internal allocation of the coalition ex ante, before the endowments of outsiders are realized, to maximize the probability that the incumbent survives. The designer can be interpreted as the leader of the coalition, or as a reduced form for selection among coalitions, in which those with more resilient allocations rule for longer.

An *allocation* $a = (p_k, x_k)_{k\in I}$ assigns positive power and resources to the members of I, with $\sum_{k\in I} p_k = P_I$ and $\sum_{k\in I} x_k = X_I$. Fix G, (P_O, X_O), and $\mathcal{Z}_O$ as in Section 4.1, and let $E_a \subseteq \mathcal{Z}_O$ denote the set of states at which the incumbent survives under a. Given a probability distribution μ on $\mathcal{Z}_O$, the designer's problem is to maximize $\mu(E_a)$ over a feasible set $\mathcal{F}$ of allocations. We assume that $\mathcal{F}$ has three properties:

(i) *Feasibility.* Assumptions 1–2 hold at every allocation in $\mathcal{F}$ and every state in $\mathcal{Z}_O$.

(ii) *Internal stability.* I is internally stable at every allocation in $\mathcal{F}$, that is, Condition (i) of Proposition 3 holds.

(iii) *Closure under exchanges.* If $a \in \mathcal{F}$ and an exchange as in Proposition 4 turns a into a', then $a' \in \mathcal{F}$.

Properties (ii) and (iii) are consistent, because an exchange preserves internal stability (see the proof of Proposition 4). Such a set exists in Example 1: the internally stable allocations in which every member holds at least a given positive amount of resources, provided that κ is small enough and that every outsider holds at least the same amount in every state. Let $\mathcal{F}^h$ denote the set of hierarchical allocations in $\mathcal{F}$.

Corollary 3 (Optimality of hierarchical allocations) *For every allocation $a \in \mathcal{F}$, there exists a hierarchical allocation $a^h \in \mathcal{F}^h$ that weakly dominates a state by state: the incumbent survives under a^h in every state in which it survives under a, that is, $E_a \subseteq E_{a^h}$. Hence, for every probability distribution μ on $\mathcal{Z}_O$,*

$$\sup_{a \in \mathcal{F}} \mu(E_a) = \sup_{a \in \mathcal{F}^h} \mu(E_a).$$

In particular, whenever an optimal allocation exists, some hierarchical allocation is optimal.

Restricting attention to hierarchical allocations is therefore without loss of optimality. The result is also distribution-free: because a^h weakly dominates a state by state, the designer does not need to know μ. If the incumbent survives under a^h but not under a with positive probability, then a is suboptimal; the example in Appendix C shows that this can occur.

Two caveats apply. First, the corollary compares an allocation only with a hierarchical allocation reached from it by exchanges. It does not rank two hierarchical allocations, or a hierarchical allocation and an arbitrary non-hierarchical one. Second, an exchange changes the payoffs of the two members involved, and we impose no participation constraint on them: the corollary characterizes the choice of a survival-maximizing designer, not the outcome of bargaining among members.

Proof of Corollary 3. Fix $a \in \mathcal{F}$. If a is hierarchical, let $a^h = a$. Otherwise, the pooling procedure of Corollary 1 gives a finite sequence of exchanges as in Proposition 4 that leads from a to a hierarchical allocation a^h. By property (iii), every allocation in this sequence belongs to $\mathcal{F}$, so $a^h \in \mathcal{F}^h$, and by property (i) every exchange in the sequence is admissible in every state.

Consider one exchange in the sequence, from an allocation b to an allocation b'. The proof of Proposition 4 uses the hypothesis that I is the unique ruling coalition in only one way: through Condition (i) of Proposition 3, which holds at b by property (ii). Its arguments therefore show that $\mathcal{S}_b^{\text{ext}} \subseteq \mathcal{S}_{b'}^{\text{ext}}$, where $\mathcal{S}_b^{\text{ext}}$ denotes the external safe area of I under the allocation b. Applying this to every exchange in the sequence gives $\mathcal{S}_a^{\text{ext}} \subseteq \mathcal{S}_{a^h}^{\text{ext}}$.

The external safe area does not depend on the state, and I is internally stable under both a and a^h by property (ii). Equation (8) therefore implies $E_a \subseteq E_{a^h}$, and hence $\mu(E_a) \leq \mu(E_{a^h})$ for every probability distribution μ on $\mathcal{Z}_O$.

Since $\mathcal{F}^h \subseteq \mathcal{F}$, we have $\sup_{a \in \mathcal{F}^h} \mu(E_a) \leq \sup_{a \in \mathcal{F}} \mu(E_a)$. The reverse inequality holds because every $a \in \mathcal{F}$ has some $a^h \in \mathcal{F}^h$ with $\mu(E_a) \leq \mu(E_{a^h})$. Finally, if a^* maximizes $\mu(E_a)$ over $\mathcal{F}$, then $\mu(E_{(a^*)^h}) \geq \mu(E_{a^*})$, so the hierarchical allocation $(a^*)^h$ is also a maximizer. ■

Proof of Lemma 2. Fix a coalition I and $\lambda > 0$, and define $A_\lambda := \{k \in I : p_k - \lambda x_k > 0\}$. Suppose first that $A_\lambda \neq \emptyset$. For any $A \subseteq I$, define $\Phi_\lambda(A) := P_A - \lambda X_A = \sum_{k \in A}(p_k - \lambda x_k)$. By construction, A_λ maximizes Φ_λ over all subsets of I. If some $B \subseteq I$ satisfied $P_B > P_{A_\lambda}$ and $X_B < X_{A_\lambda}$, then $\Phi_\lambda(B) - \Phi_\lambda(A_\lambda) = (P_B - P_{A_\lambda}) - \lambda(X_B - X_{A_\lambda}) > 0$, contradicting maximality. Hence $A_\lambda \in \mathcal{A}_I$.

For part (ii), let $i, j \in I$ be a mis-ordered pair, so $p_i > p_j$ and $x_i < x_j$. Then $\frac{p_i}{x_i} > \frac{p_j}{x_j}$. For any $\lambda \in (\frac{p_j}{x_j}, \frac{p_i}{x_i})$, we have $p_i - \lambda x_i > 0$ and $p_j - \lambda x_j < 0$. Thus $i \in A_\lambda$, $j \notin A_\lambda$, and in particular $A_\lambda \neq \emptyset$, so part (i) applies. ■

Proof of Proposition 5. Fix the pair i, j and the cutoff coalition $A^* := A_\lambda$ chosen in the proposition. By Lemma 2, $A^* \in \mathcal{A}_I$, with $i \in A^*$ and $j \notin A^*$.

Recall that $\overline{H}_I(P) = \max_{A \in \mathcal{A}_I} H_I^A(P)$ is the upper envelope of the translated boundaries. By local decisiveness, there exists $P_0 \in [0, P_O)$ such that $H_I^{A^*}(P_0) > H_I^A(P_0)$ for every $A \in \mathcal{A}_I \setminus \{A^*\}$. In particular, $H_I^{A^*}(P_0)$ is finite and $\overline{H}_I(P_0) = H_I^{A^*}(P_0)$.

The same strict inequality holds for every nonempty $A \subseteq I$ with $A \neq A^*$. Indeed, if $A \notin \mathcal{A}_I$ and its translated boundary is active at P_0, some best sub-coalition B has strictly more power and strictly fewer resources than A. Because $P_0 < P_O$, the boundary of B is also active at P_0, and $H_I^B(P_0) > H_I^A(P_0)$ because H_I is strictly increasing. Since $H_I^{A^*}(P_0) \geq H_I^B(P_0)$, we obtain $H_I^{A^*}(P_0) > H_I^A(P_0)$.

Now consider the admissible first exchange between i and j. Let $p_{i'} = p_i - \Delta p$, $x_{i'} = x_i + \Delta x$, $p_{j'} = p_j + \Delta p$, and $x_{j'} = x_j - \Delta x$, where $0 \leq \Delta p \leq \frac{p_i - p_j}{2}$, $0 \leq \Delta x \leq \frac{x_j - x_i}{2}$, and $(\Delta p, \Delta x) \neq (0, 0)$. Let I^1 denote the post-exchange coalition. The exchange preserves the aggregate power and resources of the ruling coalition: $P_{I^1} = P_I$ and $X_{I^1} = X_I$. Hence the indifference curve through the ruling coalition is unchanged.

Fix a nonempty $A' \subseteq I^1$ whose translated boundary is active at P_0, and let $A^- \subseteq I$ be its pre-exchange counterpart, constructed as in the proof of Proposition 4: the same members if A' contains both i' and j' or neither, with i in place of i' if A' contains i' but not j', and with i in place of j' if A' contains j' but not i'. That construction does not use the fact that A' is a best sub-coalition, and it gives $P_{A^-} \geq P_{A'}$ and $X_{A^-} \leq X_{A'}$. Since $P_0 < P_O$, the boundary of A^- is active at P_0, and

$$H_{I^1}^{A'}(P_0) = H_I(P_0 + P_{A'}) - X_{A'} \leq H_I(P_0 + P_{A^-}) - X_{A^-} = H_I^{A^-}(P_0).$$

If $A^- \neq A^*$, then $H_I^{A^-}(P_0) < H_I^{A^*}(P_0)$ by the previous paragraph. Because $i \in A^*$ and $j \notin A^*$, the counterpart is A^* in two cases only. If $A' = (A^* \setminus \{i\}) \cup \{i'\}$, then $P_{A'} = P_{A^*} - \Delta p$ and $X_{A'} = X_{A^*} + \Delta x$, with at least one of Δp and Δx positive. If $A' = (A^* \setminus \{i\}) \cup \{j'\}$, then $P_{A'} = P_{A^*} - (p_i - p_j - \Delta p)$ and $X_{A'} = X_{A^*} + (x_j - x_i - \Delta x)$, where $p_i - p_j - \Delta p \geq \frac{p_i - p_j}{2} > 0$ and $x_j - x_i - \Delta x \geq \frac{x_j - x_i}{2} > 0$ because i and j are mis-ordered. In both cases, $H_{I^1}^{A'}(P_0) < H_I^{A^*}(P_0)$ because H_I is strictly increasing. Inactive boundaries equal $-\infty$. Hence $H_{I^1}^{A'}(P_0) < H_I^{A^*}(P_0)$ for every nonempty $A' \subseteq I^1$, and therefore

$$\overline{H}_{I^1}(P_0) < H_I^{A^*}(P_0) = \overline{H}_I(P_0).$$

The translated boundary of the full ruling coalition is active at P_0 and equals $H_I(P_0 + P_I) - X_I \geq H_I(P_I) - X_I = 0$. Since $I \in \mathcal{A}_I$ and $I \neq A^*$, local decisiveness gives $\overline{H}_I(P_0) > 0$. Choose X such that $\max\{0, \overline{H}_{I^1}(P_0)\} < X < \overline{H}_I(P_0)$. The point (P_0, X) belongs to $\mathcal{D}_O$, is safe after the exchange, and is unsafe before it. Together with the weak inclusion in Proposition 4, this gives $\mathcal{S}_I^{\text{ext}} \subsetneq \mathcal{S}_{I^1}^{\text{ext}}$.

Starting from I^1, continue applying admissible exchanges, using the pooling procedure of Corollary 1, until the coalition reaches the hierarchical allocation I'. By Proposition 4, each subsequent exchange weakly expands the external safe area. Thus $\mathcal{S}_{I^1}^{\text{ext}} \subseteq \mathcal{S}_{I'}^{\text{ext}}$. Combining this inclusion with the strict first step gives $\mathcal{S}_I^{\text{ext}} \subsetneq \mathcal{S}_{I'}^{\text{ext}}$. Hence the sequence of exchanges leading to the hierarchical allocation strictly increases external resilience. ■

Proof of Proposition 6. Because the exchange preserves the aggregate power and resources of I, P_I and X_I are unchanged. Hence each individual indifference curve $H_k^I(\cdot)$ is unchanged.

Consider any nonempty sub-coalition of I after the exchange. We construct a pre-exchange coalition whose unanimity boundary weakly dominates its post-exchange boundary.

If the coalition contains neither i nor j, or contains both, use the identical pre-exchange coalition. Its aggregate power and resources are unchanged, so its unanimity boundary is unchanged.

Suppose next that the coalition contains i but not j, and use the same set of members before the exchange. Relative to its post-exchange counterpart, the pre-exchange coalition has Δp more power and Δx fewer resources. Whenever the post-exchange boundary is active, the corresponding pre-exchange coalition is proper, so its aggregate power is below P_N. Hence, for every member k of the coalition, $H_k^I(q + \Delta p) + \Delta x > H_k^I(q)$, where q is the post-exchange aggregate power of the deviation. Thus its pre-exchange unanimity boundary is strictly above its post-exchange boundary.

Finally, suppose that the post-exchange coalition is $A' = C \cup \{j\}$, with $i \notin A'$, and compare it with the pre-exchange coalition $A = C \cup \{i\}$. If q denotes the aggregate power of the post-exchange deviation, then the pre-exchange deviation has power $q + d_p$ and d_x fewer internal resources. Whenever the post-exchange boundary is active, this pre-exchange coalition is proper, so $q + d_p < P_N$ and Condition (12) applies. For every common member $k \in C$, we have $H_k^I(q + d_p) + d_x \geq H_k^I(q)$, because H_k^I is increasing and $d_p, d_x \geq 0$. For the exchanged member, Condition (12) gives $H_i^I(q + d_p) + d_x \geq H_j^I(q)$. Subtracting $X_{A'}$ from both sides of each inequality and taking the minimum on each side gives $\underline{H}_I^A(P) \geq \underline{H}_{I'}^{A'}(P)$ for every feasible outsider power P for which the post-exchange boundary is active. If the post-exchange boundary is inactive, the inequality needed for the safe-area comparison is automatic.

Thus every post-exchange sub-coalition has a pre-exchange counterpart whose unanimity boundary is weakly higher. Taking the maximum over all post-exchange sub-coalitions yields $\underline{H}_{I'}(P) \leq \underline{H}_I(P)$ for every feasible outsider power P. Therefore $\mathcal{S}_{I,\text{het}}^{\text{ext}} \subseteq \mathcal{S}_{I',\text{het}}^{\text{ext}}$.

If a proper winning sub-coalition A' were unanimously profitable after the exchange, then $\underline{H}_{I'}^{A'}(0) > 0$. The comparison above gives a proper winning pre-exchange sub-coalition A with $\underline{H}_I^A(0) \geq \underline{H}_{I'}^{A'}(0) > 0$, contradicting initial internal stability.

For strictness, let A^* and P_0 be as in Condition (c), and write $A^* = C \cup \{i\}$, where C contains neither i nor j. By Condition (c), $\underline{H}_I^{A^*}(P_0) > 0$. Its post-exchange counterpart contains i but not j, loses Δp units of power, and gains Δx units of resources. Its unanimity boundary therefore falls strictly at P_0. Consider any other post-exchange sub-coalition. If its pre-exchange counterpart is different from A^*, its boundary is weakly below a pre-exchange boundary that is strictly below $\underline{H}_I^{A^*}(P_0)$ by local decisiveness. The only other post-exchange coalition whose pre-exchange counterpart can be A^* is $C \cup \{j\}$. The strict initial gaps and the bounds on $(\Delta p, \Delta x)$ imply $d_p, d_x > 0$, so the common members' component boundaries are strictly lower after the exchange, and the strict version of Condition (12) gives the same strict comparison for the exchanged member. Its unanimity boundary is therefore strictly below $\underline{H}_I^{A^*}(P_0)$ as well. Hence $\underline{H}_{I'}(P_0) < \underline{H}_I(P_0)$. Choose X such that $\max\{0, \underline{H}_{I'}(P_0)\} < X < \underline{H}_I(P_0)$. The point (P_0, X) is safe after the exchange and unsafe before it. Together with weak inclusion, this gives $\mathcal{S}_{I,\text{het}}^{\text{ext}} \subsetneq \mathcal{S}_{I',\text{het}}^{\text{ext}}$. ■

Proof of Corollary 2. For any $q \in [\beta P_N, P_N - d_p)$,

$$H_i^I(q + d_p) - H_i^I(q) = \int_q^{q+d_p} (H_i^I)'(s)\, ds \geq m_i d_p.$$

Therefore $H_i^I(q+d_p)+d_x-H_j^I(q)\geq m_i d_p+d_x-D_{ji}$. Inequality (13) implies Condition (12); a strict inequality implies its strict version. The conclusions then follow from Proposition 6. If $H_i^I(q)\geq H_j^I(q)$ for every $q\in[\beta P_N, P_N-d_p)$, then $D_{ji}\leq 0$, while $d_x+m_i d_p\geq 0$, so the weak condition follows immediately. ■

Proof of Proposition 7. Consider any $W\in Z$ such that $P_W<P_I$. Since I maximizes $G(\cdot)$ over Z, $G(P_I,X_I)\geq G(P_W,X_W)$. Because $G(\cdot)$ is decreasing in internal resources, this is equivalent to $X_W\geq H_I(P_W)$. Equation (14) gives $\widetilde{H}_I(P_W)-H_I(P_W)=-\int_{P_W}^{P_I}\left[\widetilde{H}_I'(Q)-H_I'(Q)\right]dQ<0$. It follows that $X_W\geq H_I(P_W)>\widetilde{H}_I(P_W)$, and therefore $\widetilde{G}(P_I,X_I)>\widetilde{G}(P_W,X_W)$. No coalition in Z with power strictly below P_I can maximize $\widetilde{G}(\cdot)$. Hence $P_{\widetilde{I}}\geq P_I$.

Suppose now that $\widetilde{I}\neq I$ and $P_{\widetilde{I}}=P_I$. Since I maximizes $G(\cdot)$, monotonicity in internal resources implies $X_{\widetilde{I}}\geq X_I$. If the inequality is strict, then $\widetilde{G}(P_{\widetilde{I}},X_{\widetilde{I}})<\widetilde{G}(P_I,X_I)$, which contradicts the choice of $\widetilde{I}$. Equality would imply that two distinct potential ruling coalitions have the same aggregate power and resources, contrary to genericity. Therefore $P_{\widetilde{I}}>P_I$ whenever $\widetilde{I}\neq I$. ■

Proof of Proposition 8. We proceed in three steps.

Step 1. We first show that every best sub-coalition with weakly less power than C_q is non-winning. Suppose, to the contrary, that C_s is winning for some $s\geq q$. Since I is the unique ruling coalition, $G(P_I,X_I)>G(P_{C_s},X_{C_s})$, which is equivalent to $X_{C_s}>H_I(P_{C_s})$. On the other hand,

$$\begin{aligned}X_I-H_I(P_{C_s})&=H_I(P_I)-H_I(P_{C_s})\\&\leq \bar{h}(P_I-P_{C_s})\\&<X_I-X_{C_s}.\end{aligned}$$

For $s=q$, the last inequality follows from the $r=1$ term of the first minimum in (16). For $s>q$, it follows from the corresponding term of the second minimum. Hence $H_I(P_{C_s})>X_{C_s}$, a contradiction. Therefore, $P_{C_s}<\beta P_N$ for every $s\geq q$.

Step 2. We construct an outsider-power level at which the translated boundary of C_q is active while every lower-power boundary is inactive. Condition (15) is equivalent to $\beta P_N-P_{C_q}<P_O$. Since C_q is non-winning, $\beta P_N-P_{C_q}>0$. If $q<m$, choose P_0 such that $\beta P_N-P_{C_q}<P_0<\min\{\beta P_N-P_{C_{q+1}},\,P_O\}$. The interval is nonempty because $P_{C_q}>P_{C_{q+1}}$ and Condition (15) holds. If $q=m$, choose $P_0\in(\beta P_N-P_{C_q},\,P_O)$. In either case, $P_0\in[0,P_O)$, $H_I^{C_q}(P_0)$ is finite, and $H_I^{C_s}(P_0)=-\infty$ for every $s>q$.

Step 3. It remains to compare C_q with best sub-coalitions that have more power. For every $r<q$,

$$\begin{aligned}H_I^{C_r}(P_0)-H_I^{C_q}(P_0)&=H_I(P_0+P_{C_r})-H_I(P_0+P_{C_q})\\&\quad-(X_{C_r}-X_{C_q})\\&\leq \bar{h}(P_{C_r}-P_{C_q})-(X_{C_r}-X_{C_q})\\&<0,\end{aligned}$$

where the last inequality follows from (16). Thus, $H_I^{C_q}(P_0)>H_I^{C_r}(P_0)$ for every $r<q$. Together with Step 2, this shows that C_q uniquely determines the upper envelope at P_0. The same

boundaries are active in a neighborhood of P_0 by the choice in Step 2, so the strict inequalities persist by continuity.

Finally, Condition (15) is exactly the relevance requirement in Condition 1. We have also shown that C_q is locally decisive. Proposition 5 therefore applies to an admissible sequence leading to hierarchy that starts with an exchange of the chosen pair. ■

Proof of Proposition 9. We proceed in two steps.

Step 1. Consider any proper best sub-coalition $A \in \mathcal{A}_I \setminus \{I\}$ and any $P \in [0, P_O)$. If $P + P_A < \beta P_N$, the translated boundary of A is inactive. If it is active, the bound $H_I'(P) \geq \psi$ gives

$$\begin{aligned} H_I^I(P) - H_I^A(P) &= H_I(P + P_I) - H_I(P + P_A) - (X_I - X_A) \\ &\geq \psi(P_I - P_A) - (X_I - X_A) \\ &> 0, \end{aligned}$$

where the last inequality follows from $\psi > \Psi_I$. The translated boundary of I therefore lies strictly above every active translated boundary of a proper best sub-coalition. Hence $\overline{H}_I(P) = H_I(P + P_I) - X_I$.

Step 2. We show that the same boundary remains dominant after the exchange. Write $p_{i'} = p_i - \Delta p$, $x_{i'} = x_i + \Delta x$, $p_{j'} = p_j + \Delta p$, and $x_{j'} = x_j - \Delta x$, where $0 \leq \Delta p \leq \frac{p_i - p_j}{2}$, $0 \leq \Delta x \leq \frac{x_j - x_i}{2}$, and $(\Delta p, \Delta x) \neq (0, 0)$. We first establish the following comparison. For every proper $A' \in \mathcal{A}_{I'} \setminus \{I'\}$, there exists a proper $A \in \mathcal{A}_I \setminus \{I\}$ such that

$$P_A \geq P_{A'} \qquad \text{and} \qquad X_A \leq X_{A'}. \tag{21}$$

To prove the comparison, construct a pre-exchange coalition $A^- \subseteq I$. If A' contains both i' and j', replace them by i and j. If it contains neither, leave it unchanged. In either case, $P_{A^-} = P_{A'}$ and $X_{A^-} = X_{A'}$. If A' contains i' but not j', replace i' by i. Then $P_{A^-} \geq P_{A'}$ and $X_{A^-} \leq X_{A'}$. Finally, if A' contains j' but not i', replace j' by i. The bounds on the exchange imply $p_i \geq p_j + \Delta p = p_{j'}$ and $x_i \leq x_j - \Delta x = x_{j'}$, so the same two weak inequalities hold.

If $A^- \in \mathcal{A}_I$, take $A = A^-$. Otherwise, some sub-coalition of I has strictly more power and strictly fewer resources than A^-. If that sub-coalition is not best, repeat the argument. Since I is finite and power rises strictly at every step, the procedure ends at some $A \in \mathcal{A}_I$ that satisfies the inequalities in (21). Moreover, A is proper. The procedure cannot end at I, since I has strictly more resources than every proper sub-coalition, whereas each step requires strictly fewer resources.

For the pair A, A' in (21), we have $X_I - X_{A'} \leq X_I - X_A$ and $P_I - P_{A'} \geq P_I - P_A$. Since all resource differences and power differences in the ratios below are positive, $\frac{X_I - X_{A'}}{P_I - P_{A'}} \leq \frac{X_I - X_A}{P_I - P_A} \leq \Psi_I$. Thus, the lower bound $\psi > \Psi_I$ also makes the full-coalition boundary strictly dominant after the exchange. Since $P_{I'} = P_I$ and $X_{I'} = X_I$, the indifference curve through the ruling coalition is unchanged and

$$\begin{aligned} \overline{H}_{I'}(P) &= H_I(P + P_{I'}) - X_{I'} \\ &= H_I(P + P_I) - X_I \\ &= \overline{H}_I(P). \end{aligned}$$

The external safe areas therefore coincide for $P \in [0, P_O)$. The point (P_O, X_O) belongs to both safe areas by Proposition 4. Hence the external safe areas are identical. ■

Proof of Proposition 10. Since both indifference curves pass through (P_I, X_I), Condition (b) and the rotation identity (14) imply $\widetilde{H}_I(P) < H_I(P)$ for every $P \in [\beta P_N, P_I)$ and $\widetilde{H}_I(P) > H_I(P)$ for every $P \in (P_I, P_N)$.

Consider first internal resilience. For $P < \beta P_N$, sub-coalitions of I are non-winning and therefore automatically safe under both technologies. For $P \in [\beta P_N, P_I)$, internal safety under $H_I(\cdot)$ requires $X > H_I(P)$, while internal safety under $\widetilde{H}_I(\cdot)$ requires $X > \widetilde{H}_I(P)$. Since $\widetilde{H}_I(P) < H_I(P)$ on this interval, $\mathcal{S}_I^{\text{int},-}(H_I) \subseteq \mathcal{S}_I^{\text{int},-}(\widetilde{H}_I)$. If $\beta P_N < P_I$, the two positive boundaries differ for $P \in (\beta P_N, P_I)$, so a resource level strictly between them establishes strict inclusion.

Now consider external resilience. By Conditions (a) and (b), $\widetilde{H}_I'(P) > H_I'(P) \geq \psi$ for every $P \in (\beta P_N, P_N)$, so the slope-bound comparison in Step 1 of the proof of Proposition 9 applies under both technologies. Thus the full ruling coalition I determines the upper envelope on $[0, P_O)$ under both technologies: $\overline{H}_I(P) = H_I(P + P_I) - X_I$ and $\widetilde{\overline{H}}_I(P) = \widetilde{H}_I(P + P_I) - X_I$. Both boundaries equal zero at $P = 0$. For every $P \in (0, P_O)$,

$$\begin{aligned}
\widetilde{\overline{H}}_I(P) - \overline{H}_I(P) &= \widetilde{H}_I(P + P_I) - H_I(P + P_I) \\
&= \int_{P_I}^{P+P_I} \left[\widetilde{H}_I'(Q) - H_I'(Q)\right] dQ \\
&> 0.
\end{aligned}$$

On $[0, P_O)$, safety requires resources to lie strictly above the corresponding boundary. Thus $\mathcal{S}_I^{\text{ext}}(\widetilde{H}_I) \subseteq \mathcal{S}_I^{\text{ext}}(H_I)$ on this interval. Since (P_O, X_O) is safe under both technologies by Condition (c), the inclusion holds on $\mathcal{D}_O$. Because I is proper and all powers are positive, $P_O > 0$. For any $P \in (0, P_O)$, both boundaries are positive and the one under $\widetilde{H}_I(\cdot)$ is strictly higher. A resource level strictly between them establishes strict inclusion. ■

B Dynamic interpretation of the resilience analysis

The resilience analysis of Section 4 has a dynamic interpretation. Consider the model of the main text as one period of a repeated game. At the start of each period, the ruling coalition chooses its internal allocation before it observes the distribution of power and resources among outsiders. Once that distribution is realized, the static game determines whether the incumbent survives, that is, whether it remains the unique ruling coalition. For shocks that hold the outsiders' aggregate power and resources fixed, Equation (8) states when the incumbent survives in terms of the external safe area. The internal safe area describes resilience to secession within the period. A statement about the probability of surviving internal shocks would require a probability distribution over the endowments of insiders.

Two examples illustrate this interpretation. The first is a repeated environment with a king, an army, and farmers. In each period, the palace extracts resources from the countryside but does not exhaust it: the farmers survive and rebuild their resources. Exogenous shocks redistribute power and resources among the farmers in each period, while their aggregate power and resources remain fixed. The static game then determines whether some group of outsiders, possibly joined by defecting insiders, can change the membership of the ruling coalition. As shown in Section 4.1, for two internally stable allocations of the incumbent with the same aggregates, a larger external safe area weakly increases the probability that the incumbent

survives, for any probability distribution over states. The increase is strict if and only if the states in which the incumbent survives under one allocation but not under the other have positive probability.

In the second example, the plundering technology varies across periods. Suppose that the ruling coalition does not know in advance how effectively its power will translate into plunder. The realized technology then determines the indifference curves and hence the safe areas, so the conditions on the marginal return to power in Propositions 8 and 9 hold in some periods and not in others. In a period in which the marginal return to power is high, admissible exchanges leave external resilience unchanged. In a period in which it is low and a cutoff coalition is relevant and locally decisive, an admissible sequence of exchanges that begins with an exchange between the corresponding mis-ordered pair strictly increases external resilience, as in Proposition 5.

More generally, the resilience analysis applies to a dynamic environment whenever the coalition chooses a feasible internal allocation before it observes the state, and the interaction within each period starts from the incumbent and follows the game of the main text. The static model is then the one-period building block of a dynamic analysis.

C A reduced-form example satisfying Assumptions 1–3

At a generic baseline profile, we give a reduced-form example that satisfies Assumptions 1–3. The example is useful because it makes the central trade-off transparent: a ruling coalition wants power, but adding insiders also protects more resources from plunder.

Example 1 *For any $i \in I \subset N$, suppose*

$$w_i(I) = G_i(I) := \kappa \left(\frac{p_i}{P_I}\right) \left(\frac{P_I}{P_N}\right)^{\alpha+1} \left(\frac{X_N}{X_I}\right), \tag{22}$$

where $\alpha > 0$ and $\kappa > 0$. The term $\frac{p_i}{P_I}$ is player i's intra-coalition share. The term $\left(\frac{P_I}{P_N}\right)^{\alpha+1} \left(\frac{X_N}{X_I}\right)$ is the coalition-level plundering technology. It increases with aggregate coalition power and decreases with internal resources. The latter feature captures the basic plundering trade-off: resources held by insiders are protected from plunder and therefore reduce the resources available outside the coalition.

For every proper coalition I, define outsiders' plundering losses by $w_j(I) := -\frac{x_j}{X_{N\setminus I}} \sum_{i\in I} w_i(I)$ for every $j \notin I$, and set $w_i(N) = 0$ for every $i \in N$. Because the set of proper coalitions is finite, κ can be chosen small enough that $\sum_{i\in I} w_i(I) \leq X_{N\setminus I}$ for every proper I. These payoffs then satisfy the resource-feasibility requirement in Assumption 1. For any specified finite pooling sequence, choose one κ in advance small enough for every proper coalition at every step and hold it fixed along the sequence. Finiteness and positive endowments guarantee such a choice. This constructs an admissible sequence, but does not imply that any given baseline κ remains feasible after exchanges. If the example is combined with the outsider-shock exercise in Section 4.1, we restrict $\mathcal{Z}_O$ to states for which the same κ continues to satisfy this requirement.

Normalize $P_N = X_N = 1$. Then $G_i(I) = \kappa \frac{p_i P_I^\alpha}{X_I}$. Thus Assumption 2 holds with $g(i) = \kappa p_i$ and $G(P, X) = \frac{P^\alpha}{X}$. Choose α outside the finite set of values that generate a tie $\frac{P_I^\alpha}{X_I} = \frac{P_J^\alpha}{X_J}$ for distinct proper winning coalitions I and J. When $P_I \neq P_J$, each pair excludes at most the value

$\alpha = \frac{\log\left(\frac{X_I}{X_J}\right)}{\log\left(\frac{P_I}{P_J}\right)}$; *when* $P_I = P_J$, *primitive genericity implies* $X_I \neq X_J$, *so no tie is possible. Thus Assumption 3 holds. Fix a plundering gain* $\overline{G}_i > 0$. *Player* i*'s indifference curve at this gain consists of the pairs* (P, X) *satisfying* $G_i(P, X) = \overline{G}_i$, *or*

$$X = C_i P^{\alpha}, \tag{23}$$

where $C_i := \frac{\kappa p_i}{\overline{G}_i}$ *and* $P \in [\beta, 1)$. *Along this curve,* $\frac{dX}{dP} = \frac{\alpha X}{P}$. *Thus* α *governs the slope of the power–resource trade-off. For curves through the same incumbent,* $H_\alpha(P) = X_I \left(\frac{P}{P_I}\right)^{\alpha}$ *and* $H'_\alpha(P) = \frac{\alpha X_I}{P_I}\left(\frac{P}{P_I}\right)^{\alpha-1}$. *The change in this slope with* α *has the sign of* $\frac{1}{\alpha} + \log\left(\frac{P}{P_I}\right)$, *so increasing* α *need not give uniformly higher slopes below* P_I. *The uniform slope conditions in the main text must be checked separately.*

The economic interpretation follows directly. At fixed X, *the elasticity of* $G(P, X)$ *with respect to power is* α. *At a given* (P, X), *the power needed to compensate for a marginal increase in protected resources is* $dP = \frac{P}{\alpha X} dX$, *which falls as* α *rises.*

The curvature of the curve in (23) *is also immediate. If* $0 < \alpha < 1$, *indifference curves are strictly concave. If* $\alpha = 1$, *they are linear. If* $\alpha > 1$, *they are strictly convex. Thus the common ranking function* $G(P, X) = \frac{P^\alpha}{X}$ *is concave or convex in power according to* α. *The coalition-level extraction function is proportional to* $\frac{P^{\alpha+1}}{X}$ *and is convex in power for every* $\alpha > 0$.

An example with a low marginal return to power. Set $\alpha = 1$, so that $G(P, X) = \frac{P}{X}$ and indifference curves are linear. Set $\beta = 0.6$ and $\kappa = 10^{-3}$, with

$$p = (0.11, 0.16, 0.44, 0.22, 0.07),$$
$$x = (0.07, 0.46, 0.05, 0.36, 0.06).$$

The unique ruling coalition is $I = \{1, 3, 5\}$, with $P_I = 0.62$ and $X_I = 0.18$; its value $G(I) \approx 3.4444$ exceeds the next-highest winning value, approximately 1.6098. The indifference curve through (P_I, X_I) is $H_I(P) = \frac{0.18}{0.62} P$, with constant slope $\bar{h} = \frac{9}{31} \approx 0.2903$.

The best sub-coalitions of I, in order of decreasing power, are $C_1 = \{1, 3, 5\}$, $C_2 = \{1, 3\}$, $C_3 = \{3, 5\}$, and $C_4 = \{3\}$, with aggregate power and resources $(0.62, 0.18)$, $(0.55, 0.12)$, $(0.51, 0.11)$, and $(0.44, 0.05)$. Players 3 and 1 are mis-ordered: $p_3 = 0.44 > p_1 = 0.11$ and $x_3 = 0.05 < x_1 = 0.07$. The power-to-resource ratios of players 3, 1, and 5 are 8.8, approximately 1.571, and approximately 1.167, so the cutoff coalition is $A_\lambda = \{3\} = C_4$ for every $\lambda \in \left(\frac{p_1}{x_1}, \frac{p_3}{x_3}\right)$. It is relevant, because $P_{C_4} = 0.44 > P_I - (1 - \beta)P_N = 0.22$. The three ratios in the first minimum of (16) are $\frac{0.13}{0.18} \approx 0.722$, $\frac{0.07}{0.11} \approx 0.636$, and $\frac{0.06}{0.07} \approx 0.857$, and the second minimum is over an empty set. Condition (16) therefore holds: $\bar{h} \approx 0.290 < 0.636$. The marginal return to power is low, and Proposition 8 implies that $\{3\}$ is locally decisive. Indeed, at outsider power $P_0 = 0.17$, the translated boundary of $\{3\}$ is $\bar{h} \times 0.61 - 0.05 \approx 0.1271$, whereas those of $\{1, 3\}$, $\{3, 5\}$, and I are approximately 0.0890, 0.0874, and 0.0494.

The pooling procedure takes a single step. Players 3 and 1 are pooled, and each receives power 0.275 and resources 0.06. The transfers are $\Delta p = 0.165$ and $\Delta x = 0.01$, so the pooling is a single exchange between players 3 and 1 as in Proposition 4. Player 5, with power 0.07 and resources 0.06, is not mis-ordered with either of them. The result is a hierarchical allocation with two ranks, $\{1, 3\}$ and $\{5\}$; Condition (20) fails at $k = 2$. After pooling, the same set of

members is the unique ruling coalition, and at $P_0 = 0.17$ the upper envelope is the translated boundary of $\{1, 3\}$, approximately 0.0890. At this level of outsider power, the upper envelope therefore falls from approximately 0.1271 to approximately 0.0890, as Proposition 5 implies.

Let outsider shocks satisfy $p_2 = p$, $p_4 = 0.38 - p$, $x_2 = y$, and $x_4 = 0.82 - y$, on the common domain $p \in (0.05, 0.33)$ and $y \in (0.05, 0.77)$. Every player's resources are at least 0.05 throughout this domain and the exchange path. For every proper coalition C, total plunder is $\frac{\kappa P_C^2}{X_C}$; its ratio to available outsider resources is at most $\frac{\kappa}{0.05 \times 0.05} = 0.4$. Thus the same value of κ satisfies resource feasibility for every proper coalition throughout the exchange sequence.

For every state with $p \in (0.165, 0.175)$ and $y \in (0.095, 0.120)$, the threshold that the cutoff coalition $\{3\}$ sets for player 2 before pooling exceeds 0.1256, and player 2's threshold after pooling is below 0.0905. Player 4's threshold after pooling is below 0.1021, while her resources exceed 0.70; (P_O, X_O) is safe under both allocations. These states are therefore unsafe before pooling and safe afterward. Under the uniform distribution on the common domain, hierarchy strictly increases survival probability.