The Optimal Degree of Centralization

Charles Z. Zheng*

April 10, 2016

Abstract

This paper considers a normative foundation of government based on a Hobbesian anarchy where one region may pillage the other. A center that takes enough power from the regions and helps the victim to fight the aggressor can deter warfare thereby incentivizing productive efforts of the regions, but with less power the regions become less productive. This paper provides a method to calculate the social-surplus maximizing allocation of power. When the marginal product of power and complementarity between power and efforts are sufficiently small, every fully decentralized system generates less social surplus than a system that centralizes power to some degree. When the regions are identical in their parameters, among the centralized systems that treat them equally, the optimal degree of centralization decreases as productive capability expands. In a numerical example, all fully decentralized systems are outperformed by even the total centralization, which strips all power from the regions and maximizes tax revenues extracted from them to the center.

*Department of Economics, University of Western Ontario, London, Ontario, charles.zheng@uwo.ca, http://economics.uwo.ca/faculty/zheng/.

1 Introduction

Should the authority of a society be decentralized or centralized? How much authority should be centralized, if at all? These questions are fundamental to a society or organization, whether it is the crowd of migrants led by Moses in the *Exodus* and *Leviticus*, or the Persians upon overthrowing the Magi usurpation, with Darius and his comrades debating the choice of government forms among democracy, oligarchy and monarchy,[1] or a corporate conglomerate choosing between expanding through franchisees or company-owned units, or the United States citizens debating the right to bear arms amidst recurring gun violence, or even the current international community, with territorial disputes, cultural clashes and a possibly looming climate crisis. In addressing these questions the Enlightenment philosophers disagree, ranging from Hobbes's [21] proposal of total centralization to Rousseau's [43] insistence on direct democracy. Theoretical literature in economics and political economy has addressed such questions through binary comparisons between two specific games, mostly within a democratic setting, one corresponding to decentralization and the other, centralization. Yet these questions are seldom addressed for more general, not necessarily democratic, environments. And the binary comparisons have yet to provide a method to locate an optimal degree of centralization in the entire spectrum between the two extremes.

To fill the gap, this paper falls back to the primitive setting based on which Enlightenment philosophers such as Locke [31] and especially Hobbes [21] consider the normative foundation of government. This is the state of nature, modeled here as a noncooperative game between two regions, each allocated some power, who may pillage each other with the odds of winning proportional to their power. If at the outset an enough amount of power has been transferred to a center that will join force with the victim to fight the aggressor, then at equilibrium aggression is deterred, property rights secured, and the regions exert more productive efforts. The catch is that the regions, with less power, become less productive. This paper characterizes the social-surplus maximum among all allocations of power based on their equilibrium outcomes (Theorems 1 and 2). The power allocated to the center at such optimums is the optimal degree of centralization.

Rather than the one-size-fits-all solutions proposed by Hobbes and Rousseau, the optimal allocation of power, depending on the parameters, may be fully decentralized, leaving

[1] Herodotus [20, III, 80–86].

all power to the regions, or centralized, moving some power to the center (Remark 2). Yet we also have general patterns. Pointing to a normative foundation for centralization, Theorem 4 essentially says that if the marginal product of power and the complementarity between power and efforts are sufficiently small then every fully decentralized system is outperformed by some centralized one. When the two regions are identical in their parameters, Theorem 3 says that the optimal degree of centralization, among the centralized systems that allocate equal power to the regions, decreases in the productive capability of the regions.

While the focus of the paper is the normative foundation of centralization, with the main results based on the assumption of a neutral, trustworthy center, the paper also considers cases where the center is greedy, maximizing the revenue extracted from the regions. If such a greedy center can commit to a tax rate then it prefers to commit to a rate that is not draconian (Propositions 5 and 6). Moreover, in a numerical example, any fully decentralized system is outperformed by even the worst among centralized systems, the total centralization that strips all power from the regions, with the greedy, absolutely powerful center committing to a tax rate to maximize its own revenues (Remark 3).

These results are relevant to historical questions why some societies came to be more centralized than others. For example, a question in comparative history is why the Roman empire was never restored after her fall in 476 C.E. whereas the Chinese empire was reunified again and again.[2] And it has been suggested that China's backwardness in recent centuries was due to her centralized institutions.[3] The results of this paper would suggest a different idea, that such different degrees of centralization between two societies could be ascribed to their best responses to their different environments.[4] For instance, in the spirit of Theorem 4, the power allocated to the local levels might be less important to production activities in the land-based, partially landlocked China than in the Mediterranean world, prone to shocks from all directions and calling for quick responses at the local levels. A similar comparison could also be made between the land-based France and the sea-based Britain,[5]

[2] Scheidel [44] draws an interesting parallel between the two empires up to the fifth century and raises the question why afterwards the two diverged to opposite directions, one almost fully decentralized, and the other mostly centralized. More of such works are cited in http://web.stanford.edu/~scheidel/Divergence.pdf.

[3] For example, Mokyr [35] and Myerson [39].

[4] A pioneer of such perspective in the study of history is Jared Diamond [16].

[5] Observations that France was more centralized than Britain date back at least to a Byzantine historian Chalkokondyles at the final period of Byzantium, whose contrast of France and Britain is summarized in Gibbon [17, v6, pp403–405].

or between the land-based Sparta with kings and helots versus the sea-based Athens in direct democracy.[6] For another instance, Theorem 3 points to productive capability as a fundamental driving force for decentralization. It is no accident that the industrial revolution was roughly contemporary with democratization of Britain.

The model of the paper may be applicable to organization theory regarding the optimal management of internal competition.[7] There, our pillage game would correspond to the territorial encroachment between franchisees of the same brand,[8] or the cannibalization between the brick-and-mortar and e-commerce branches of a retail chain,[9] or the competition for readership between the print and digital news groups of a media company,[10] or infringement of implicit property rights of ideas between two divisions vying to be designated as the developer of a frontier technology.[11] Balancing the innovation and initiative benefits of such decentralized structures against the ensuing cannibalization costs, corporate executives strive to find an optimal degree of such internal competition. Similar tradeoffs have also been observed in the literature of corporate finance.[12]

Our pillage game is a model of situations where involved parties have potentially adversarial relationships. Such situations have been studied from different angles in the literature: the coalitional pillage games in Jordan [24, 25, 26] and Polemarchakis and Rowat [41] through a cooperative method, as well as a noncooperative formulation by Acemoglu, Egorov and Sonin [2], to explain how an allocation of power remains stable despite lack of commit-

[6] The sea- versus land-power contrast between Athens and Sparta is apparent in Thucydides [46]. For their different degrees of centralization, a telling episode is that, when Aristagoras approached both cities with a proposal to invade Persia, it was a king who listened to the proposal and made the decision for Sparta, whereas for Athens, it was a crowd of "thirty thousand Athenians" (Herodotus [20, V, 49–52, 98]).

[7] See Birkinshaw [12] for a preview of such topics.

[8] See Kalnins [27] for an empirical study of such territorial encroachment among franchisees.

[9] The price reduction effect of online competition has been formalized by Bakos [6] with a model of differentiated products. Hedlund and Frick [19] present a case study of the internal competition between brick-and-mortar and e-commerce branches within a retail chain.

[10] *The Atlantic* is an early example of a news media encouraging aggressive competition between its own print and digital divisions, according to a news story in *The New York Times*, Dec. 13, 2010, B1.

[11] According to Birkinshaw [12], when the headquarter of Hewlett-Packard encouraged its Toronto division to continue its project to develop X terminal workstation, the California division "started crying foul and complaining that Toronto was stealing their charter" (p24).

[12] The competition between two firms has a cannibalistic effect on the portfolio of the shareholders who have invested on both firms, as observed and analyzed in Kraus and Rubin [30].

ment; the resource wars in Acemoglu, Golosov, Tsyvinski and Yared [3], Caselli, Morelli and Rohner [15] and Morelli and Rohner [37], on the relationship between natural resource endowments and wars; the literature on the reason for wars such as Baliga and Sjötröm [8] and Jackson and Morelli [22], recently surveyed by Baliga and Sjötröm [9] and Jackson and Morelli [23], to explain why wars happen given the premise that wars are costly exogenously. This paper studies warring situations with a different focus: how internal warfare can be reduced by centralization of power and whether such centralization is worthwhile with endogenous costs of war versus peace.

While the pillage game in our model may be extreme with respect to the interregional relationships within a stable country, the different interests across regions can become so contentious that render the model relevant. In fact, the theme of this paper is consistent with an idea of captures in the fiscal federalism literature, dating back to James Madison's *Federalist Papers* (no. 10), that local governments, left alone, may be captured by locally powerful interest groups and hence should be counterbalanced by the central authority of the federal government. This idea, formalized by Bardhan and Mookherjee [10] in a model of election in both the national and local levels, has given rise to a literature in political economy that compares a decentralized game with a centralized one: Besley and Coate [11], between a system where different districts make their own decisions that have spillovers to other districts and a system where an elected central legislature decides for all districts; Blanchard and Shleifer [13], between a system where local governments choose their own policies to capture rents and a system where the central government has some control of the local governments; Cai and Treisman [14], who compare a pair of systems similar to that in Blanchard and Shleifer with an additional effect that the rents captured by local governments may weaken the central government; Ponce-Rodríguez, Hankla, Martinez-Vazquez and Heredia-Ortiz [42], who compare four games that differ from one another in the degrees of government and party centralization. Focusing on a different aspect of decentralization than the above works, Myerson [38] compares between a federal voting mechanism that allows politicians to build their reputations at equilibrium and a unitary one that does not.

In the economics literature on decentralization,[13] the conflict of interests across players

[13] The academic interest in decentralization dates back at least to Hayek [18], who suggests that the price system is essentially a decentralized system to process diffused, incomplete information, which the Walrasian price theory he says has missed to capture. Pushing Hayek's point further, Stiglitz [45] criticizes

is usually modeled as an externality or coordination problem. Much of the theoretical works is again binary comparisons between a decentralized game and a centralized one: Acemoglu, Aghion, LeLarge, van Reenen and Zilibotti [1], who compare delegated management with centralized control of technology adoption within a firm; Alonso, Dessein and Matouschek [5], between a two-player cheap-talk game where the two players make decisions independently, and a direct mechanism where the center makes the decisions on their behalf without money transfer; Baliga and Sjötröm [7], between a principal-agent game where the wage depends only on the agents' output, and a more centralized one where the wage depends on an informed agent's report about the other agent's action; Klibanoff and Poitevin [29], between a bargaining game of two regions with externalities and a direct revelation mechanism without money transfer;[14] Maskin, Qian and Xu [32], between a multi-divisional authority structure ("M form") and a unitary one ("U form").

This paper differs from these literature in considering the entire spectrum between full decentralization and total centralization and proposing a model to calculate the optimums in this spectrum. While the model is relatively abstract in institutional details, our observation of the opposite welfare consequences rendered by a selfish center, with commitment versus without, suggests the need for institutional monitoring in the center.

the neoclassical or Warasian paradigm for failing to distinguish itself from market socialism, a model of centralized economies run by planners using Walrasian prices as instruments; for an alternative that might articulate the advantage of decentralization Stiglitz suggests asymmetric information. However, if one casts the design of political institutions into a problem of designing revelation mechanisms to alleviate the friction of asymmetric information, it would appear that decentralization can never outperform centralization, since by the revelation principle any equilibrium outcome of any communication mechanism can be replicated by a direct revelation mechanism, which taken literally would be centralized. Consequently, theoretical research on decentralization has developed into two strands. One is to find conditions under which outcomes of the optimal direct revelation mechanism (optimal centralization) can be implemented through decentralized mechanisms. Much of the literature on delegation belongs to this strand, on which Mookherjee [36] has a recent survey. The other strand, such as those cited here, is to restrict permissible mechanisms to a smaller set, usually containing only two elements, and compare the performance of centralization versus decentralization within this subset.

[14] Klibanoff and Morduch [28] consider a coordination problem in a decentralized model of externalities.

2 The Model

A society consists of two players, region 1 and region 2, as well as a center, which is assumed neutral, and not a player, until Section 7. The society is endowed with a resource, called *power*, and its quantity is normalized to one, perfectly divisible. First, a *power allocation* $(x_1, x_2) \in [0,1]^2$ such that $x_1 + x_2 \leq 1$ is determined so that each region $i \in \{1,2\}$ gets the quantity x_i of power, with the remaining $1 - x_1 - x_2$ kept by the center. Second, each region i chooses an effort level $e_i \in \mathbb{R}_+$ thereby producing a quantity $f_i(x_i, e_i)$ of output and bearing a sunk cost equal to $c_i e_i$, with $c_i > 0$ a parameter and $f_i : [0,1] \times \mathbb{R}_+ \to \mathbb{R}_+$. Third, given output quantities (y_1, y_2) produced by the two regions, each region i decides, simultaneously, whether to pillage the other region, denoted $-i$. If i pillages $-i$ while $-i$ does not pillage i, i wins with probability

$$\frac{x_i}{1 - x_1 - x_2 + x_{-i} + x_i} = x_i,$$

meaning that the center joins force with $-i$ to fight i; if both regions pillage the other, then the center stays neutral and each region i's probability of winning equals $\frac{x_i}{x_{-i}+x_i}$. In case of internal warfare, i.e., that at least one region pillages the other, the winning region gets the entire output $y_1 + y_2$ of both, and the losing side, zero. A region's payoff is equal to the quantity of output it obtains in the end minus the sunk cost of the effort exerted by the region; and each region is assumed risk-neutral.

We are interested in the equilibrium outcomes of various power allocations, with **equilibrium** being any pair of efforts exerted by the regions, coupled with their pillage plans contingent on the ensuing outputs, that constitutes a subgame perfect equilibrium, given the power allocation, without weakly dominated strategies. A power allocation (x_1, x_2) is said *fully decentralized* if $x_1 + x_2 = 1$. Otherwise, the allocation is *centralized* more or less, with the degree of centralization measured by $1 - x_1 - x_2$, the quantity of power kept, unproductively, by the center. A *system*, or allocation-equilibrium pair, is *peaceful* if internal warfare occurs with zero probability at equilibrium, and *warring* if otherwise.

The following **assumptions** are made on the function f_i of each region i, with $D_k g$ denoting the partial derivative $\frac{\partial}{\partial z_k} g$ with respect to the kth variable of any differentiable multivariate function g, and $D_k^2 g := D_k D_k g$: f_i is strictly increasing in each argument; f_i is twice continuously differentiable over $\mathbb{R}_+^2$; for any $x_i \in [0,1]$, $\lim_{e_i \to \infty} D_2 f_i(x_i, e_i) < c_i < \lim_{e_i \downarrow 0} D_2 f_i(x_i, e_i)$; $D_2^2 f_i < 0$ and $D_1 D_2 f_i \geq 0$.

Interpretation: We can think of power as a durable resource that helps a region in both production and wars. From the standpoint of an empire, a region's "productive" activity can be a foreign war against peripheral peoples, and its output the ensuing wealth. Power in this context corresponds to the weaponry and logistics in the region and the authority delegated to the regional governors; naturally these factors improve the odds for the region to prevail in wars, foreign or internal. More generally, we can think of a region i as a productive unit, and its power x_i the discretion and logistic support that the unit has, so that the power contributes an input $g(x_i)$ to the unit's production, and a competitive edge $h(x_i)$ in the rivalry with the other production unit such that the odds of winning are proportional to $h(x_i)$. When h is additive, such situations can be turned into the model presented above, with $h(x_i)$ playing the role of x_i, and $f_i\left(g\left(h^{-1}(\hat{x}_i)\right), e_i\right)$ that of $f_i(x_i, e_i)$.

3 Preliminary Analysis

3.1 The Pillage Game Given Realized Outputs

Let us start by considering the endgame, when a power allocation (x_1, x_2) has been given, and outputs (y_1, y_2) already produced.

Lemma 1 *Given any $y_i \in \mathbb{R}_+$ as the quantity of region i's output, with $i \in \{1, 2\}$:*

a. if the other region is to pillage region i then i weakly prefers to stay put (i.e., to not pillage $-i$), and strictly prefers so if $x_1 + x_2 < 1$;

b. if $x_i(y_i + y_{-i}) > y_i$ then region i strictly prefers to pillage the other region $-i$ when $-i$ does not pillage i; if $x_i(y_i + y_{-i}) \leq y_i$ then region i weakly prefers to stay put, and strictly prefers so if this inequality is strict.

Proof Given outputs (y_1, y_2) of the two regions, the payoff matrix of the endgame is:

	pillage	not pillage
pillage	$\frac{x_1}{x_1+x_2}(y_1+y_2), \frac{x_2}{x_1+x_2}(y_1+y_2)$	$x_1(y_1+y_2), (1-x_1)(y_1+y_2)$
not pillage	$(1-x_2)(y_1+y_2), x_2(y_1+y_2)$	y_1, y_2

The lemma follows directly from this payoff matrix. $\blacksquare$

Lemma 2 *For each region i, any power allocation (x_1, x_2) and any output pair $(y_1, y_2) \in \mathbb{R}^2_+$, if $x_i(y_i + y_{-i}) \geq y_i$ then $x_{-i}(y_i + y_{-i}) \leq y_{-i}$, and the last inequality is strict if either $x_i(y_i + y_{-i}) > y_i$ or $x_1 + x_2 < 1$.*

Proof Note the equivalence between $x_i(y_i + y_{-i}) \geq y_i$ and $x_i y_{-i} + (1 - x_i) y_{-i} \geq (1 - x_i)(y_i + y_{-i})$, as well as the equivalence between the two when the weak inequalities are replaced by strict ones. Since $1 - x_i \geq x_{-i}$, the lemma follows. ■

Lemma 3 *At any allocation-equilibrium pair (x_1, x_2, e_1, e_2), if a region totally mixes the actions between pillaging the other region and not doing so, then $x_1 + x_2 = 1$.*

Proof Given any allocation-equilibrium pair (x_1, x_2, e_1, e_2), denote $y_i := f_i(x_i, e_i)$ for each $i \in \{1, 2\}$. Suppose that $x_1 + x_2 < 1$ and region 1 totally mixes between "pillage" and "not pillage." Then by the payoff matrix in the proof of Lemma 1, we have $x_1(y_1 + y_2) > y_1$, otherwise "pillage" is weakly dominated. By Lemma 2, this inequality implies $x_2(y_1 + y_2) < y_2$, which in turn implies that "not pillage" is the unique best response for region 2 (Lemma 1.b), but then at equilibrium region 1's best response is uniquely to pillage $-i$ rather than totally mix between pillaging and not pillaging, a contradiction. ■

3.2 A Region's Expected Payoff from Its Effort

Anticipating the pillage game analyzed above and expecting the quantity y_{-i} of output that the other region is going to produce, each region i can calculate its gross payoff as a function of the effort it exerts, as in Eq. (1). The top branch there corresponds to the case where region i expects to be the aggressor at the endgame, the bottom branch the case where i expects to be the victim, and in the middle one neither region will pillage the other.

Lemma 4 *Given any power allocation $x := (x_1, x_2)$, for each region i and any $(e_i, y_{-i}) \in \mathbb{R}^2_+$, if region i expects the other region to produce a quantity y_{-i} of output, then i's expected payoff from exerting effort e_i is equal to $f_i^*(x, e_i, y_{-i}) - c_i e_i$, where*

$$f_i^*(x, e_i, y_{-i}) := \begin{cases} x_i\left(f_i(x_i, e_i) + y_{-i}\right) & \text{if } (1 - x_i) f_i(x_i, e_i) \leq x_i y_{-i} \\ f_i(x_i, e_i) & \text{if the following (2) holds} \\ (1 - x_{-i})\left(f_i(x_i, e_i) + y_{-i}\right) & \text{if } x_{-i} f_i(x_i, e_i) \geq (1 - x_{-i}) y_{-i}; \end{cases} \tag{1}$$

in addition, not pillaging the other region is a best response for both regions if and only if

$$(1 - x_i) f_i(x_i, e_i) \geq x_i y_{-i} \quad \text{and} \quad x_{-i} f_i(x_i, e_i) \leq (1 - x_{-i}) y_{-i}. \tag{2}$$

Proof Appendix A. ■

A region's gross expected payoff $f_i^*(x, e_i, y_{-i})$ characterized above, as a function of e_i, is depicted in Figure 1, where the cutoffs e_i^L and e_i^H are shorthands for

$$e_i^L(x, y_{-i}) := \inf\{e_i \in \mathbb{R}_+ : (1-x_i)f_i(x_i, e_i) \geq x_i y_{-i}\}, \tag{3}$$

$$e_i^H(x, y_{-i}) := \sup\{e_i \in \mathbb{R}_+ : x_{-i}f_i(x_i, e_i) \leq (1-x_{-i}) y_{-i}\}. \tag{4}$$

These cutoffs are determined by the anticipated output y_{-i} of the other region. For region i, exerting an effort less than $e_i^L(x, y_{-i})$ means that it plans to pillage $-i$ in the endgame, exerting effort above $e_i^H(x, y_{-i})$ means that i expects to be pillaged by $-i$, and any effort between the two cutoffs means peace. Such observation is formalized by the next lemma.

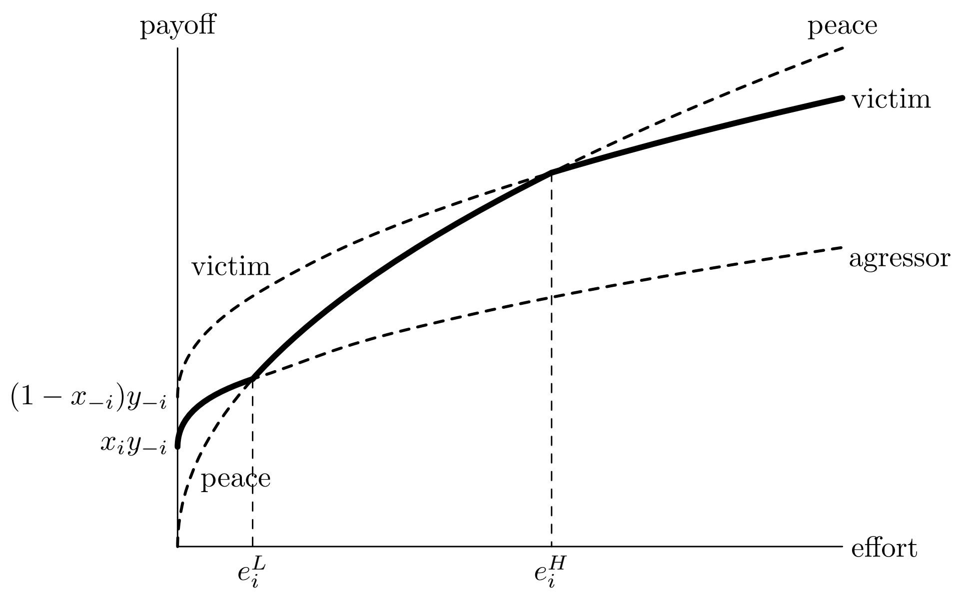

Figure 1: The gross expected payoff for region i

Lemma 5 *For any region i, power allocation $x := (x_1, x_2)$, $e_i \in \mathbb{R}_+$ and $y_{-i} \in \mathbb{R}_{++}$, there exist thresholds $e_i^L(x, y_{-i})$ and $e_i^H(x, y_{-i})$, with $e_i^L(x, y_{-i}) \leq e_i^H(x, y_{-i})$, such that*

$$f_i^*(x, e_i, y_{-i}) = \begin{cases} x_i\left(f_i(x_i, e_i) + y_{-i}\right) & \text{if } e_i \leq e_i^L(x, y_{-i}) \\ f_i(x_i, e_i) & \text{if } e_i^L(x, y_{-i}) \leq e_i \leq e_i^H(x, y_{-i}) \\ (1-x_{-i})\left(f_i(x_i, e_i) + y_{-i}\right) & \text{if } e_i \geq e_i^H(x, y_{-i}) \end{cases} \tag{5}$$

and the following properties are true:

a. $e_i <$ *(resp.,* $\leq, >$*)* $e_i^L(x, y_{-i}) \iff x_i\left(f_i(x_i, e_i) + y_{-i}\right) >$ *(resp.,* $\geq, <$*)* $f_i(x_i, e_i)$*;*

b. $e_i < (\textit{resp.}, \leq, >)\ e_i^H(x, y_{-i}) \Longleftrightarrow (1 - x_{-i})\,(f_i(x_i, e_i) + y_{-i}) > (\textit{resp.}, \geq, <)\ f_i(x_i, e_i)$;

c. $f_i^*(x, \cdot, y_{-i})$ *is strictly concave on* $\left[0, e_i^L(x, y_{-i})\right]$*, with a jump in its derivative at the point* $e_i^L(x, y_{-i})$ *if and only if* $e_i^L(x, y_{-i}) < e_i^H(x, y_{-i})$*, and is again strictly concave on* $\left[e_i^L(x, y_{-i}), \infty\right)$*;*

d. $x_1 + x_2 < 1 \Longleftrightarrow e_i^L(x, y_{-i}) < e_i^H(x, y_{-i})$*;*

e. $e_i < (\textit{resp.}, \leq, =)\ e_i^L(x, y_{-i}) \Longleftrightarrow e_{-i} > (\textit{resp.}, \geq, =)\ e_{-i}^H(x, y_i)$.

Proof Appendix A. ■

3.3 Effort Levels That Are Possible Best Replies

A region has only two kinds of possible best responses, one being an optimal effort level when it anticipates peace, and the other optimal when it anticipates war. Figure 2 depicts one of such cases, where the supporting hyperplane at a peace-anticipating effort level $\psi_i(x_i)$ happens to be higher than the one at a pillage-planning effort level $\eta_i(x_i)$, so the region's best response is $\psi_i(x_i)$. However, even a small change in the power allocation (x_1, x_2) or the other region's anticipated output y_{-i} may alter the relative position between the two supporting hyperplanes, switching the region's best response discontinuously from a peace-anticipating effort level to a war-anticipating one. Following are useful facts of the comparative statics of both kinds of possible best responses, which we shall see are special cases of a function $\gamma_i : [0,1]^2 \to \mathbb{R}_+$ defined by, for any $(x_i, z_i) \in [0,1]^2$,

$$\{\gamma_i(x_i, z_i)\} := \arg\max_{e_i \in \mathbb{R}_+} z_i f_i(x_i, e_i) - c_i e_i. \tag{6}$$

Lemma 6 *For each* $i \in \{1, 2\}$*:*

a. for any $(x_i, z_i) \in [0,1]^2$*,* $\gamma_i(x_i, z_i)$ *defined by Eq. (6) exists and is unique;*

b. for any $x_i \in [0,1]$ *there exists* $\zeta_i(x_i) \in [0,1)$ *such that if* $z_i \leq \zeta_i(x_i)$ *then* $\gamma_i(x_i, z_i) = 0$*, and if* $z_i > \zeta_i(x_i)$ *then* $\gamma_i(x_i, z_i) > 0$ *and* $z_i D_2 f_i\left(x_i, \gamma_i(x_i, z_i)\right) = c_i$*;*

c. $\gamma_i(x_i, z_i)$ *is weakly increasing in* (x_i, z_i)*, and strictly increasing in* z_i *if* $z_i > \zeta_i(x_i)$*;*

d. γ_i *is continuous on* $[0, 1[^2$.

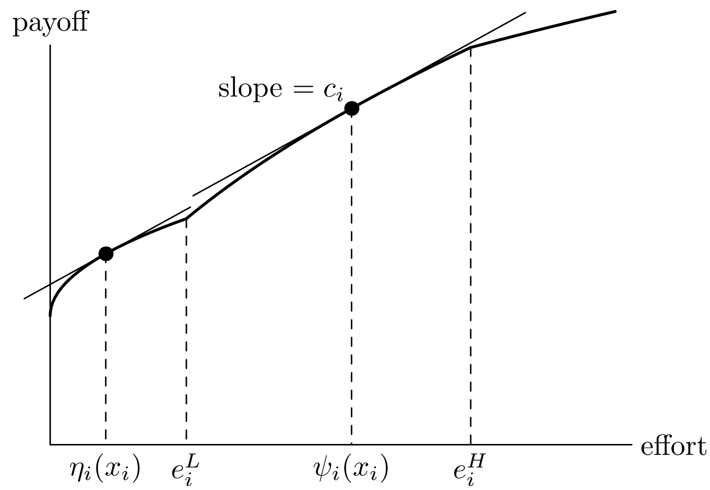

Figure 2: $\psi_i(x_i) = \gamma_i(x_i, 1)$; $\eta_i(x_i) = \gamma_i(x_i, x_i)$

Proof Appendix A. ■

Lemma 6 applied to the cases $z = 1$ and $z = x_i$ gives Lemma 7, whose proof we omit.

Lemma 7 *For each $i \in \{1, 2\}$ and any $x_i \in [0, 1]$ there exist a unique $\psi_i(x_i) \in \mathbb{R}_{++}$ and a unique $\eta_i(x_i) \in \mathbb{R}_+$ such that, for any $y_{-i} \in \mathbb{R}_+$,*

$$\{\psi_i(x_i)\} = \arg\max_{e_i \in \mathbb{R}_+} f_i(x_i, e_i) - c_i e_i, \tag{7}$$

$$\{\eta_i(x_i)\} = \arg\max_{e_i \in \mathbb{R}_+} x_i \left(f_i(x_i, e_i) + y_{-i}\right) - c_i e_i, \tag{8}$$

$$c_i = D_2 f_i(x_i, \psi_i(x_i)), \tag{9}$$

and ψ_i and η_i are each continuous.

For each region $i \in \{1, 2\}$, the following inequalities are true for any $x_i \in [0, 1)$, and the equalities true for any $x_i \in [0, 1]$ and any $y_{-i} \in \mathbb{R}_+$. They are proved in Appendix A.

$$\psi_i(x_i) > \eta_i(x_i), \tag{10}$$

$$f_i\left(x_i, \psi_i(x_i)\right) - c_i \psi_i(x_i) > f_i\left(x_i, \eta_i(x_i)\right) - c_i \eta_i(x_i), \tag{11}$$

$$\frac{d}{dx_i}\psi_i(x_i) = -\frac{D_1 D_2 f_i(x_i, \psi_i(x_i))}{D_2^2 f_i(x_i, \psi_i(x_i))}, \tag{12}$$

$$\frac{d}{dx_i}\left(f_i\left(x_i, \psi_i(x_i)\right) - c_i \psi_i(x_i)\right) = D_1 f_i(x_i, \psi_i(x_i)) > 0. \tag{13}$$

4 Decentralization

To characterize the social-surplus maximums among all possible power allocations, we dissect the entire family of allocation-equilibrium pairs into two, one consisting of all warring systems, and the other all peaceful systems. Our tractable characterization of the first set is due to an observation that, as far as social surplus concerns, warring systems are equivalent to fully decentralized ones.

4.1 Full Decentralization Means War

A consequence of full decentralization, giving all power to the regions, is internal warfare. During the production phase, as formalized in the next lemma and proposition, any effort level that a region chooses to exert will render the region either an aggressor or victim in the endgame, except for the threshold effort level at which the region is indifferent between war and peace. But the threshold effort level becomes a region's best reply only when the region's decision is equivalent to the case as if it is expecting war.

Lemma 8 *Given any fully decentralized power allocation $x := (x_1, x_2)$ and any quantity $y_{-i} \in \mathbb{R}_+$ of output that region $-i$ will produce, $f_i^*(x, \cdot, y_{-i}) = x_i(f_i(x_i, \cdot) + y_{-i})$ on $\mathbb{R}_+$.*

Proof Since $x_1 + x_2 = 1$ by definition of full decentralization, the top and bottom branches of Eq. (1) coincide. To complete the proof, pick any e_i for which the middle branch applies, i.e., when (2) holds. With $x_1 + x_2 = 1$, (2) is reduced to $(1 - x_i)f_i(x_i, e_i) = x_i y_{-i}$, i.e.,

$$f_i(x_i, e_i) = x_i(f_i(x_i, e_i) + y_{-i}) = (1 - x_{-i})(f_i(x_i, e_i) + y_{-i}).$$

Hence at any such e_i Eq. (1) implies that all of its three branches coincide. ■

The next proposition is attributed to Thomas Hobbes [21] for his bleak picture of the state of nature, or full decentralization in our model.[15]

Proposition 1 (Hobbes) *Any fully decentralized power allocation (x_1, x_2) admits a warring equilibrium and admits no other equilibrium with a different outcome in effort or output.*

[15] "[D]uring the time men live without a common power to keep them all in awe, they are in that condition which is called war, and such a war as is of every man against every man" (Hobbes [21, Chapter 13, p389]).

Proof Since (x_1, x_2) is fully decentralized, Lemma 8 implies that each region i's effort decision is the one in Eq. (8), which by Lemma 7 has a unique solution $\eta_i(x_i)$, which in turn determines uniquely the output of region i. That proves the uniqueness claim. To prove existence of a warring equilibrium, note from Lemma 5.d that $e_i^L(x, y_{-i}) = e_i^H(x, y_{-i})$ for any $y_{-i} \in \mathbb{R}_+$ and any $i \in \{1, 2\}$. Thus, for each i and any $e_i \in \mathbb{R}_+$, either $e_i \leq e_i^L(x, y_{-i})$ or $e_i \geq e_i^H(x, y_{-i})$. Therefore, by Lemma 2 and Eqs. (3)–(4), for any $(e_1, e_2) \in \mathbb{R}_+^2$, with ensuing outputs (y_1, y_2), there is an $i \in \{1, 2\}$ for whom $(1 - x_i)y_i \leq x_i y_{-i}$ and $x_{-i}y_i \geq (1 - x_{-i})y_{-i}$. Hence it is an equilibrium in the endgame that region $-i$ pillages region i, and i stays put (the payoff matrix at Lemma 1). ■

4.2 War, at Best, Means Full Decentralization

Other than full decentralization, there is a myriad of systems, with various degrees of centralization, some resulting in peace, and others in war. The next proposition reduces the complexity by asserting that among the alternatives of full decentralization we need only to consider those that do not necessarily result in wars. Coupled with Lemma 3, this proposition implies that in maximizing social surplus among warring systems we need only to consider fully decentralized ones.

Proposition 2 *For any allocation-equilibrium pair (x_1, x_2, e_1, e_2) such that $x_1 + x_2 < 1$ and war occurs for sure, there exists another allocation-equilibrium pair (x_1', x_2', e_1', e_2') such that $x_1' + x_2' = 1$ and it generates a larger social surplus.*

Proof Let (x_1, x_2, e_1, e_2) be a warring system, with $x_1 + x_2 < 1$ and say region 1 pillaging region 2 at equilibrium. Since $x_1 + x_2 < 1$, $e_i^L(x, f_{-i}(x_{-i}, e_{-i})) < e_i^H(x, f_{-i}(x_{-i}, e_{-i}))$ for each $i \in \{1, 2\}$ (Lemma 5.d); as region 1 is the aggressor, $e_1 < e_1^L(x, f_2(x_2, e_2))$, which by Lemma 5.e implies $e_2 > e_2^H(x, f_1(x_1, e_1))$. Thus, the effort e_1 of the aggressor solves the problem in Eq. (8) and so $e_1 = \eta_1(x_1)$ by Lemma 7; and the effort e_2 of region 2 solves

$$\max_{e_2'' \in \mathbb{R}_+} (1 - x_1) \left(f_2(x_2, e_2'') + y_1 \right) - c_2 e_2'', \tag{14}$$

which implies $e_2 = \gamma_2(x_2, 1 - x_1)$ by Eq. (6).

Consider a fully decentralized power allocation (x_1', x_2') such that $x_1' := x_1$ and $x_2' := 1 - x_1$ (hence $> x_2$). With $x_1' + x_2' = 1$, Lemma 8 implies that the expected payoff function for each region $i \in \{1,2\}$ becomes the differentiable and strictly concave function $e_i'' \mapsto x_i'\left(f_i(x_i', e_i'') + y_{-i}'\right) - c_i e_i''$ (given any y_{-i}'). That means, for region 1, the objective function becomes $x_1'\left(f_i(x_1', e_1'') + y_2'\right) - c_1 e_1'' = x_1\left(f_i(x_1, e_1'') + y_2'\right) - c_1 e_1''$ since $x_1' = x_1$, hence its decision problem is again equivalent to the one in Eq. (8), as in the previous power allocation (x_1, x_2). Thus, e_1 remains to be region 1's best response under (x_1', x_2').

Now that region 1's output remains unchanged as before, say denoted by y_1, region 2's expected payoff from exerting effort e_2'' becomes

$$x_2'(f_2(x_2', e_2'') + y_1) - c_2 e_2'' = (1 - x_1)(f_2(x_2', e_2'') + y_1) - c_2 e_2'',$$

as $x_2' = 1 - x_1$. Thus, given allocation (x_1', x_2') region 2's best response $e_2' = \gamma_2(x_2', 1 - x_1)$ by Eq. (6). Since $\gamma_2(\cdot, 1 - x_1)$ is weakly increasing (Lemma 6.c), $e_2' = \gamma_2(x_2', 1 - x_1) \geq \gamma_2(x_2, 1 - x_1) = e_2$. Thus, with $x_2' > x_2$ and f_2 assumed strictly increasing,

$$f_2(x_2', e_2') > f_2(x_2, e_2). \tag{15}$$

With $\gamma_2(x_2', \cdot)$ weakly increasing (Lemma 6.c), $e_2' = \gamma_2(x_2', 1 - x_1) \leq \gamma_2(x_2', 1)$. Hence $\gamma_2(x_2', 1) \geq e_2' \geq e_2$. This, coupled with the fact that $\gamma_2(x_2', 1)$ is the unique maximum (Lemma 6.a) of the strictly concave function $e_2'' \mapsto f_2(x_2', e_2'') - c_2 e_2''$, implies that

$$e_2' > e_2 \Rightarrow f_2(x_2', e_2') - c_2 e_2' > f_2(x_2', e_2) - c_2 e_2.$$

Consequently, since $e_2' \geq e_2$ and $f_2(\cdot, e_2)$ is strictly increasing,

$$e_2' \neq e_2 \Longrightarrow f_2(x_2', e_2') - c_2 e_2' > f_2(x_2, e_2) - c_2 e_2. \tag{16}$$

Eqs. (15) and (16), combined with the fact $(x_1', e_1') = (x_1, e_1)$, imply the desired conclusion

$$f_1(x_1', e_1') - c_1 e_1' + f_2(x_2', e_2') - c_2 e_2' > f_1(x_1, e_1) - c_1 e_1 + f_2(x_2, e_2) - c_2 e_2. \quad \blacksquare$$

4.3 Optimal Decentralization

By Propositions 1 and 2 and Lemma 3, maximizing social surplus among warring systems is equivalent to maximizing social surplus among fully decentralized ones, which is a relatively simple maximization problem with only a unidimensional choice variable.

Theorem 1 *A system that maximizes the social surplus among warring allocation-equilibrium pairs exists and is also the social surplus maximum among fully decentralized systems; it is a power allocation* $(x_1^*, 1-x_1^*)$ *such that* x_1^* *solves the problem*

$$\max_{x_1\in[0,1]} f_1(x_1,\eta_1(x_1)) - c_1\eta_1(x_1) + f_2\left(1-x_1,\eta_2(1-x_1)\right) - c_2\eta_2(1-x_1). \tag{17}$$

Proof By Propositions 1 and 2 and Lemma 3, any optimum among the warring allocation-equilibrium pairs is the solution of

$$\begin{aligned}
\max_{(x_1,x_2;e_1,e_2)\in[0,1]^2\times\mathbb{R}_+^2} \quad & \sum_{i=1}^{2}\left(f_i(x_i,e_i) - c_ie_i\right) \\
\text{subject to} \quad & x_1 + x_2 = 1, \\
& \forall i\in\{1,2\} : e_i \in \arg\max_{e_i'\in\mathbb{R}_+}\left(x_i\left(f_i(x_i,e_i') + f_{-i}(x_{-i},e_{-i})\right) - c_ie_i'\right),
\end{aligned}$$

with the second constraint equivalent to the equilibrium condition because a region's payoff function given a fully decentralized power allocation collapses to the one in the above constraint due to Lemma 8. By Eq. (8), this constraint is the same as $e_i = \eta_i(x_i)$ for each $i\in\{1,2\}$. Thus, the optimization problem is equivalent to (17). With f_i continuous by assumption and η_i continuous by Lemma 7, the solution for this problem exists. ■

5 Centralization

By Theorem 1, to seek better alternatives to warring systems, and better alternatives to full decentralization, we need only to consider peaceful systems that are not fully decentralized.

5.1 Characterization of Peaceful Centralized Systems

For peace to prevail in the endgame, each region must, during the production phase, find it a best response to exert an amount of effort such that the ensuing output preempts the pillage incentive. Obviously such an effort level needs to maximize a region's expected payoff provided that peace will prevail, i.e., to solve the problem in (7). That however does not suffice a best response for the region, because its expected payoff is not a concave function of its effort (Lemma 5.c). The next proposition characterizes the condition for peace as Ineq. (18), which is essentially a region's incentive feasibility condition of not deviating to another effort level with a plan to pillage the other region in the endgame.

Proposition 3 *A power allocation* (x_1, x_2) *admits a peaceful equilibrium and is not fully decentralized if and only if, for each* $i \in \{1, 2\}$,

$$f_i\left(x_i, \psi_i(x_i)\right) - c_i\psi_i(x_i) \geq x_i\left(f_i\left(x_i, \eta_i(x_i)\right) + f_{-i}\left(x_{-i}, \psi_{-i}(x_{-i})\right)\right) - c_i\eta_i(x_i). \tag{18}$$

Proof To prove the "only if" part, note that if (e_1, e_2) constitutes a peaceful equilibrium given allocation (x_1, x_2) then each region i's equilibrium effort e_i belongs to the interval on which i's gross payoff function $f_i^*\left(x_1, x_2, \cdot, f_{-i}(x_{-i}, e_{-i})\right) = f_i(x_i, \cdot)$. If, in addition, (x_1, x_2) is not fully decentralized, i.e, $x_1 + x_2 < 1$, then this interval is nondegenerate (Lemma 5.d). Then e_i is a maximum of the function $e_i' \mapsto f_i(x_i, e_i') - c_ie_i'$ on a nondegenerate interval and hence, with the function concave, on the entire $\mathbb{R}_+$. Thus $e_i = \psi_i(x_i)$ by Eq. (7). Thus, if (18) is not satisfied then region i would deviate to exert effort $\eta_i(x_i)$ and then pillage $-i$.

To prove the "if" part, we first prove that (18) implies that (x_1, x_2) is not fully decentralized. To prove that, recall from Eq. (8) the fact that $\eta_i(x_i)$ maximizes

$$x_i\left(f_i\left(x_i, e_i'\right) + f_{-i}\left(x_{-i}, \psi_{-i}(x_{-i})\right)\right) - c_ie_i'$$

among all $e_i' \in \mathbb{R}_+$. Hence (18) implies

$$\begin{aligned} f_i\left(x_i, \psi_i(x_i)\right) - c_i\psi_i(x_i) &\geq x_i\left(f_i\left(x_i, \eta_i(x_i)\right) + f_{-i}\left(x_{-i}, \psi_{-i}(x_{-i})\right)\right) - c_i\eta_i(x_i) \\ &\geq x_i\left(f_i\left(x_i, \psi_i(x_i)\right) + f_{-i}\left(x_{-i}, \psi_{-i}(x_{-i})\right)\right) - c_i\psi_i(x_i). \end{aligned} \tag{19}$$

Analogously we obtain an inequality for $-i$. Summing these two inequalities we have

$$\begin{aligned} & f_1\left(x_1, \psi_1(x_1)\right) - c_1\psi_1(x_1) + f_2\left(x_2, \psi_2(x_2)\right) - c_2\psi_2(x_2) \\ \geq\ & x_1\left(f_1\left(x_1, \eta_1(x_1)\right) + f_2\left(x_2, \psi_2(x_2)\right)\right) - c_1\eta_1(x_1) + x_2\left(f_2\left(x_i, \eta_2(x_2)\right) + f_1\left(x_1, \psi_1(x_1)\right)\right) - c_2\eta_2(x_2) \\ \geq\ & x_1\left(f_1\left(x_1, \psi_1(x_1)\right) + f_2\left(x_2, \psi_2(x_2)\right)\right) - c_1\psi_1(x_1) + x_2\left(f_2\left(x_2, \psi_2(x_2)\right) + f_1\left(x_1, \psi_1(x_1)\right)\right) - c_2\psi_2(x_2). \end{aligned}$$

If the allocation (x_1, x_2) is fully decentralized, i.e., $x_1 + x_2 = 1$, then the last line of the above formula is equal to the first line. Consequently, for each $i \in \{1, 2\}$,

$$x_i\left(f_i\left(x_i, \eta_i(x_i)\right) + f_{-i}\left(x_{-i}, \psi_{-i}(x_{-i})\right)\right) - c_i\eta_i(x_i) = x_i\left(f_i\left(x_i, \psi_i(x_i)\right) + f_{-i}\left(x_{-i}, \psi_{-i}(x_{-i})\right)\right) - c_i\psi_i(x_i),$$

which implies, with Ineq. (10) true for all $x_i \in [0, 1)$, that $x_i = 1$. But that cannot be a feasible power allocation for both regions i, a contradiction.

To complete the proof for the "if" part, let $e_i := \psi_i(x_i)$ for each i and we claim that (e_1, e_2) constitutes a peaceful equilibrium given (x_1, x_2). We first note that, given the outputs produced by such effort levels, with $y_j := f_j(x_j, e_j)$ for each $j \in \{1,2\}$, each region finds it a best response to not pillage the other. That is because, for each $i \in \{1,2\}$, Ineq. (19) implies

$$y_i - x_i(y_i + y_{-i}) \geq c_i\left(\psi_i(x_i) - \psi_i(x_i)\right) = 0,$$

which in turn implies, by Lemma 1.b, that i weakly prefers to not pillage $-i$. Next we show for each i that e_i maximizes region i's expected payoff $f_i^*\left(x_1, x_2, e_i'', y_{-i}\right) - c_i e_i''$ among all $e_i'' \in \mathbb{R}_+$. First, since it is a best response for each region to not pillage the other given (x_1, x_2, e_1, e_2), Lemma 4 implies for each i that e_i belongs to the domain on which $f_i^*(x, \cdot, y_{-i}) = f_i(x_i, \cdot)$; this domain, by Eq. (5), is the interval $\left[e_i^L\left(x_1, x_2, y_{-i}\right), e_i^H\left(x_1, x_2, y_{-i}\right)\right]$. Consequently, e_i, by its definition $e_i := \psi_i(x_i)$ and Eq. (7), is a maximum of $f_i^*(x, e_i'', y_{-i}) - c_i e_i''$ among all e_i'' in this interval. The optimality of e_i is then extended to the interval $\left[e_i^H\left(x_1, x_2, y_{-i}\right), \infty\right)$, because the derivative of $f_i^*(x, \cdot, y_{-i})$ drops at $e_i^H\left(x_1, x_2, y_{-i}\right)$ and then keeps decreasing on this interval (Lemma 5.c). Thus, the optimality of e_i is further extended to the entire $\mathbb{R}_+$ due to Ineq. (18) and the definition of $\eta_i(x_i)$. Hence e_i best replies e_{-i}, as claimed. ■

Lemma 9 *For any $i \in \{1,2\}$, if $x_i = 0$ then Ineq. (18) holds strictly.*

Proof With $x_i = 0$, the right-hand side of (18) is always nonpositive. The left-hand side, by contrast, is always strictly positive: Instead of exerting effort $\psi_i(x_i)$, region i could have chosen zero effort thereby ensuring a net payoff $f_i(x_i, 0) - c_i \cdot 0 \geq 0$; but the fact that $\psi_i(x_i) > 0$ (Lemma 7) and that it is the unique solution for (7) implies $f_i(x_i, \psi_i(x_i)) - c_i\psi_i(x_i) > f_i(x_i, 0) \geq 0$. Thus, Ineq. (18) holds strictly, as claimed. ■

Corollary 1 (possibility of peace) *There exists a power allocation, not fully decentralized, that admits a peaceful equilibrium.*

Proof By Lemma 9 and Proposition 3, $(0,0)$ is such a power allocation. ■

5.2 Optimal Centralization

While peaceful equilibrium is achievable through some centralized allocation (Corollary 1), centralization would not do much good if the best it could offer is the totally centralized allocation $(0,0)$ in the proof of Corollary 1, which, like Hobbes's [21] Leviathan, maintains internal peace through depriving the regions of any power that could have helped the production. This subsection is hence devoted to characterization of the social-surplus optimum among peaceful systems that are not fully decentralized. As a corollary of the characterization, not only can other centralized systems do better than the Leviathan, total centralization, such Leviathan is also the worst among all peaceful, centralized systems. Furthermore, optimal centralization systems can be bounded away from total centralization even when power allocated to the regions is arbitrarily insignificant to production.

Lemma 10 *A system that maximizes the social surplus among peaceful allocation-equilibrium pairs that are not fully decentralized corresponds to a power allocation* (x_1^*, x_2^*) *that solves*

$$\begin{aligned} \max_{(x_1,x_2)\in[0,1]^2} \quad & \sum_{i=1}^{2} \left(f_i(x_i, \psi_i(x_i) - c_i\psi_i(x_i)\right) \\ \text{subject to} \quad & x_1 + x_2 \leq 1 \\ & (18) \text{ holds } \forall i \in \{1,2\}. \end{aligned} \tag{20}$$

Proof The constraint $x_1+x_2 \leq 1$ is merely the feasibility condition for (x_1, x_2) to be a power allocation. By Proposition 3, the constraint that (18) holds for each region i is the sufficient and necessary condition for a (feasible) power allocation to constitute a peaceful system that is not fully decentralized. Given that the allocation (x_1, x_2) is not fully decentralized, there is a nondegenerate interval on which a region's payoff function $f_i^*(x, \cdot, y_{-i})$ is equal to $f_i(x_i, \cdot)$ (Lemma 5.d). Thus, by Eq. (7), $\psi_i(x_i)$ is equal to region i's equilibrium effort level, and hence the effort level in the objective function. ∎

The next theorem asserts existence of a social-surplus maximum among peaceful systems that are not fully decentralized. That might sound surprising, as the qualification "not fully decentralized," $x_1 + x_2 < 1$, sounds as if the feasibility set for the maximization problem (20) were an open set. Actually it is a closed set because, by Proposition 3, any $(x_1, x_2) \in [0,1]^2$ that satisfies (18) for each i also satisfies the strict inequality $x_1 + x_2 < 1$.

Theorem 2 *There exists a social-surplus maximum* $(x_i, \psi_i(x_i))_{i=1}^2$ *among all peaceful allocation-equilibrium pairs that are not fully decentralized, and* $(x_i)_{i=1}^2$ *can only be:*

i. a solution of the following equation for each $i \in \{1,2\}$,

$$f_i(x_i, \psi_i(x_i)) - c_i\psi_i(x_i) = x_i\left(f_i(x_i, \eta_i(x_i)) + f_{-i}(x_{-i}, \psi_{-i}(x_{-i}))\right) - c_i\eta_i(x_i), \tag{21}$$

ii. or a solution of Eq. (21) for some $i \in \{1,2\}$ *and* (x_i, x_{-i}) *is a tangent point between the graph generated by Eq. (21) and an indifference curve of the objective function (20),* *iii. or* $x_{-i} = 0$ *for some* $i \in \{1, 2\}$ *and* $(x_i, 0)$ *is a solution of Eq. (21).*

Proof Appendix B. ■

The Hobbesian Leviathan is the worst centralized system:

Corollary 2 *Total centralization,* $(0, 0)$*, is not an optimum among peaceful systems that are not fully decentralized. Furthermore, any other peaceful systems that is not fully decentralized generates a larger social surplus than total centralization.*

Proof By Theorem 2, at any optimal centralized system Eq. (21) for some $i \in \{1, 2\}$ holds, whereas by Lemma 9 neither holds at $(0, 0)$. Hence $(0, 0)$ is suboptimal. For the second half of the corollary, pick any other peaceful systems $(x_i, \psi_i(x_i))_{i=1}^2$ that is not fully decentralized. Then $x_i \geq 0$ for both i and $x_i > 0$ for at least one i. By the inequality in (13), the social surplus generated by $(x_i, \psi_i(x_i))_{i=1}^2$ is larger than that generated by $(0, 0, \psi_1(0), \psi_2(0))$. ■

While total centralization is suboptimal, can it be nearly optimal under some parameter values? Casting a doubt on such possibility is the next corollary of Theorem 2, giving a lower bound of the total regional power at an optimum.

Corollary 3 *For each* $i \in \{1, 2\}$ *there exists a strictly positive constant*

$$\rho_i := \frac{f_i(0, \psi_i(0)) - c_i \psi_i(0)}{f_i(1, \psi_i(1) + f_{-i}(1, \psi_{-i}(1))} \tag{22}$$

such that, for any optimal centralized allocation (x_1, x_2)*,* $x_i > \rho_i$ *for at least one* $i \in \{1, 2\}$*, and the inequality holds for both* i *if* (x_1, x_2) *belongs to Case (i) in Theorem 2.*

Proof Appendix B. ■

The lower bound of regional power given in Corollary 3 can be bounded away from zero even when regional power becomes arbitrarily less significant than efforts to the region's production, in terms of the technical rate of substitution between power and efforts:

Remark 1 *There exists a sequence* $(f_1^n, f_2^n)_{i=1}^{\infty}$ *of environments such that the lower bound in Corollary 3 is bounded away from zero while the technical rate of substitution*

$$\lim_{n \to \infty} \frac{D_1 f_i^n(x_i, e_i)}{-D_2 f_i^n(x_i, e_i)} = 0.$$

for each $i \in \{1, 2\}$ *and for any* $(x_i, e_i) \in [0, 1] \times \mathbb{R}_+$*.*

Proof Appendix B. ■

5.3 Productivity and Decentralization

Within this subsection, let us specialize to a symmetric environment such that

$$f_1 = f_2 = f \quad \text{and} \quad c_1 = c_2 = c \tag{23}$$

for some function f and effort cost c satisfying the assumptions in Section 2. To see how the optimal degree of centralization may vary with the productive capability of a society, let us parameterize the latter by an index $t \in T$, for some compact interval T, such that the output produced by a region i, with power x_i and effort e_i, is equal to $f(x_i, e_i; t)$ with

$$D_3 f > 0,\ D_3 D_1 f \geq 0 \text{ and } D_3 D_2 f \geq 0. \tag{24}$$

Within this subsection we consider only *symmetric* allocations, where the two regions, identical in their primitives, are allocated an identical amount of power. Both regions given the production function $f(\cdot, \cdot; t)$, let $\tilde{x}(t)$ denote any social-surplus maximizer among all symmetric allocations that are peaceful and not fully decentralized. For any $x \in [0, 1/2]$, let $\psi(x, t)$ and $\eta(x, t)$ denote, respectively, the $\psi_i(x_i)$ and $\eta_i(x_i)$ defined in Eqs. (7)–(8), with the production function f_i there being $f(\cdot, \cdot; t)$, and x_i being x. For any continuous function $g : [0, 1/2] \times \mathbb{R}_+ \times T \to \mathbb{R}$, denote

$$\|g\|_{\sup} := \sup_{(x,e,t) \in [0,1/2] \times [\eta(x,t), \psi(x,t)] \times T} g\,(x, e, t)$$

and likewise for $\|g\|_{\inf}$.

Theorem 3 *If both (23) and (24) are satisfied and if*

$$\frac{c}{\| - D_2^2 f\|_{\inf}} \|D_3 D_2 f\|_{\sup} \leq \|\psi - \eta\|_{\inf} \|D_3 D_2 f\|_{\inf}, \tag{25}$$

then $t'' > t'$ implies $\tilde{x}(t'') > \tilde{x}(t')$.

Proof Appendix C ■

The intuition for the theorem is that the additional output of a region due to productivity improvements is fully owned by the region itself only if there is no internal warfare. Thus, expansions in productive capabilities may strengthen a region's preference of peace over war, thereby relaxing its incentive constraint and hence allowing more power to be delegated to the region without upsetting peace. The only countervailing effect is that the productivity improvement may enlarge the other region's output by so much that this region would rather shirk in its production and then pillage the enriched neighbor. Ineq. (25) assumed in the theorem is to ensure that such countervailing effect is dominated. The inequality requires that D_3D_2f, the complementarity between efforts and the productivity shock, vary sufficiently little and the marginal product of effort, D_2f, diminish sufficiently quickly as effort increases. For example, consider a case where (x, e) is additively separable from the productivity shock t, with t interpreted as an increase in natural endowments for the society such as a newfound colony. It is perhaps no accident that the "discovery" of America occurred merely decades before the Reformation, dissecting the Christendom in Western Europe.

6 The Optimality of Centralization

Since our model is even-handed on the pros and cons of centralization versus decentralization, it is natural that, when we compare the optimum among centralized systems with that among decentralized ones, the outcome depends on the values of the parameters.

Remark 2 *Suppose the two regions have an identical effort cost $c > 0$ and $f_1 = f_2 = f$ for some common production function f parameterized by $\alpha \geq 0$, $\beta \geq 0$, $\theta > 0$ and $\tau \in (0, 1)$.*[16]

a. If f is defined by, for any $x_i \in [0, 1]$ and any $e_i \in \mathbb{R}_+$,

$$f(x_i, e_i) := \sqrt{(x_i + \alpha)e_i}, \tag{26}$$

the optimal decentralized system generates larger social surplus than the optimal centralized system when $\alpha = 0$, and the opposite is true when $\alpha > 3 - 2\sqrt{2}$.

b. If f is defined by, for any $x_i \in [0, 1]$ and any $e_i \in \mathbb{R}_+$,

$$f(x_i, e_i) := (\theta x_i)^\beta \, e_i^\tau, \tag{27}$$

the optimal decentralized system yields larger social surplus than the optimal centralized system when $\beta + \tau \geq 1$, and the opposite is true when $\beta/(1 - \tau)$ is sufficiently small.

[16] Like Cobb-Douglas functions, the production functions in this remark satisfy all assumptions stated in Section 2 except the Inada condition $\lim_{e_i \downarrow 0} D_2 f(0, e_i) > c$ at $x_i = 0$.

Proof Appendix D. ■

To provide a normative foundation of centralization, the next Theorem 4 implies that, when the marginal product of power and the complementarity between power and efforts become sufficiently small, some degree of centralization generates larger social surplus than any fully decentralized system.

To illustrate the intuition, suppose that a fully decentralized power allocation (x_1^*, x_2^*) is replaced by an allocation $(\bar{x}_1, \bar{x}_2)$ that sustains peace and is not fully decentralized. The net gain thereof in the social surplus generated from region $i \in \{1, 2\}$ is equal to

$$\Delta_i(\bar{x}_i, x_i^*) := f_i\left(\bar{x}_i, \psi_i(\bar{x}_i)\right) - c_i\psi_i(\bar{x}_i) - f_i\left(x_i^*, \eta_i(x_i^*)\right) + c_i\eta_i(x_i^*). \tag{28}$$

To analyze this difference, recall the function $\gamma_i(x_i, z_i)$ defined in Eq. (6) and the fact that $\psi_i(x_i) = \gamma_i(x_i, 1)$ and $\eta_i(x_i) = \gamma_i(x_i, x_i)$. Thus,

$$\Delta_i(\bar{x}_i, x_i^*) = \Delta_i^s(\bar{x}_i, x_i^*) + \Delta_i^p(\bar{x}_i, x_i^*), \tag{29}$$

where

$$\begin{aligned}
\Delta_i^s(\bar{x}_i, x_i^*) &:= f_i\left(\bar{x}_i, \gamma_i(\bar{x}_i, 1)\right) - c_i\gamma_i(\bar{x}_i, 1) - f_i\left(\bar{x}_i, \gamma_i(\bar{x}_i, x_i^*)\right) + c_i\gamma_i(\bar{x}_i, x_i^*), \\
\Delta_i^p(\bar{x}_i, x_i^*) &:= f_i\left(\bar{x}_i, \gamma_i(\bar{x}_i, x_i^*)\right) - c_i\gamma_i(\bar{x}_i, x_i^*) - f_i\left(x_i^*, \gamma_i(x_i^*, x_i^*)\right) + c_i\gamma_i(x_i^*, x_i^*),
\end{aligned}$$

with $\Delta_i^s(\bar{x}_i, x_i^*)$ measuring the *security effect* of increasing the probability, with which region i keeps its own output, from x_i^* to one, and $\Delta_i^p(\bar{x}_i, x_i^*)$ the *disempowerment effect* of changing the power allocated to region i from x_i^* to $\bar{x}_i$. Unless $x_i^* \leq \bar{x}_i$, the two effects are opposite to each other, and the centralized system $(\bar{x}_1, \bar{x}_2)$ outperforms the decentralized (x_1^*, x_2^*) if the security effect outweighs the disempowerment effect. One readily sees that the former effect is large when the marginal product of efforts is large with the effort level ranging from $\gamma_i(\bar{x}_i, x_i^*)$ to $\gamma_i(\bar{x}_i, 1)$, while the latter effect is large in absolute value when the marginal product of power, due to the direct effect $D_1 f_i$ of power and its indirect effect $D_1 D_2 f_i$ on the marginal product of efforts, is large with the amount of power ranging from x_i^* to $\bar{x}_i$. Thus, when $D_1 f_i$ and $D_1 D_2 f_i$ are uniformly small enough, the disempowerment effect is outweighed by the security effect, and decentralization outperformed by centralization.

Theorem 4 *For any sequence $(f_1^n, f_2^n)_{n=1}^{\infty}$ of environments such that, for each n and each i, $\lim_{e_i \downarrow 0} D_2 f_i^n(x_i, e_i) = \infty$ for any $x_i \in (0, 1]$, if for each $i \in \{1, 2\}$ there exist $\bar{r}_i > 0$ and*

$b_i > 0$ *such that, with* ρ_i^n *defined by (22) and* γ_i^n *defined by (6),*

$$\liminf_{n\to\infty} \rho_i^n > \bar{r}_i, \tag{30}$$

$$\lim_{n\to\infty} \max_{(x_i,z_i)\in[\min\{\bar{r}_i,1/2\},1]^2} D_1 D_2 f_i^n\left(x_i, \gamma_i^n(x_i,z_i)\right) = 0, \tag{31}$$

$$\lim_{n\to\infty} \max_{(x_i,z_i)\in[\min\{\bar{r}_i,1/2\},1]^2} D_1 f_i^n\left(x_i, \gamma_i^n(x_i,z_i)\right)\left|D_2^2 f_i^n\left(x_i, \gamma_i^n(x_i,z_i)\right)\right| = 0, \tag{32}$$

$$\lim_{n\to\infty} \frac{\max_{(x_i,z_i)\in[\min\{\bar{r}_i,1/2\},1]^2}\left|D_2^2 f_i^n\left(x_i, \gamma_i^n(x_i,z_i)\right)\right|}{\min_{(x_i,z_i)\in[\min\{\bar{r}_i,1/2\},1]^2}\left|D_2^2 f_i^n\left(x_i, \gamma_i^n(x_i,z_i)\right)\right|} < b_i, \tag{33}$$

then, for all sufficiently large n*, the maximum social surplus among peaceful centralized systems given* (f_1^n, f_2^n) *is strictly larger than the maximum social surplus among fully decentralized ones given* (f_1^n, f_2^n)*.*

Proof Appendix E. ■

In Theorem 4, the Inada condition of $f_i(x_i,\cdot)$ for all $x_i > 0$ ensures that a region's optimal efforts such as $\eta_i(x_i)$ and $\gamma_i(x_i,z_i)$ be interior solutions and hence differentiable for all $x_i > 0$. The hypothesis (30), coupled with Corollary 3, guarantees that at the optimal centralized system the power in at least one region is bounded away from zero, the possible singularity point for a region's optimal effort. The other hypotheses concern the marginal effect of power versus efforts, with the effort level ranging along the path γ_i^n determined by the primitives. Eq. (31) requires that the complementarity between power and efforts be sufficiently small. Eq. (32) requires that the marginal product of power and the decay rate of the marginal product of efforts be sufficiently orthogonal to each other; in particular, the equation is satisfied if the marginal product of power is sufficiently small. Ineq. (33) imposes an upper bound on the variation of the decay rate of the marginal product of efforts.

All hypotheses of the theorem are satisfied by the example defined by Eq. (26) in Remark 2 when $\alpha > 0$, with n being the α there. For the example defined by Eq. (27) to satisfy all the hypotheses, we need only to amend its definition by translating the production function by a positive constant so that $f(0,\psi(0)) > 0$. One can then verify all the hypotheses, with $1/(1-\tau)$ and $(1-\tau)/\beta$ going to infinity when $n \to \infty$.[17]

[17] Let us illustrate with the function f_i defined by Eq. (26) for any $\alpha > 0$. It satisfies the Inada condition because $D_2 f_i(x_i,e_i) = (1/2)\sqrt{(x_i+\alpha)/e_i}$ and $\alpha > 0$. The proof of Remark 1 has verified Ineq. (30), with $\bar{r}_i \in (0, c_{-i}/(2(c_1+c_2)))$. To verify (31)–(33), note that $\gamma_i(x_i,z_i) = (x_i+\alpha)z_i^2/(4c_i^2)$. Hence $D_1 f_i(x_i,\gamma_i(x_i,z_i)) = z_i/(4c_i)$, $D_1 D_2 f_i(x_i,\gamma_i(x_i,z_i)) = c_i/\left(2z_i(x_i+\alpha)\right)$, $D_2^2 f_i\left(x_i,\gamma_i(x_i,z_i)\right) =$

7 Greedy Center

While Theorem 4 has demonstrated a normative justfication to keep some power away from the productive regions thereby maintaining internal peace, the center with such power might exploit its advantage to extort outputs from the regions. We shall modify the model to contrast a case where the center has no commitment power to another case where it has. Whereas the relative merit of centralization is gone in the former, not all is lost in the latter.

7.1 A Case of an Extortive Center

Let us modify the model so that, when internal warfare does not occur, with power allocation (x_1, x_2) and realized outputs (y_1, y_2), the center chooses an $s \in [0, 1]$ and demands each region to transfers a share s of its output to the center. If both regions disobey the order, a war breaks out between their joint force and the center, and the center wins with probability $1 - x_1 - x_2$; if the center wins then the center gets all the outputs, $y_1 + y_2$, else each region retains its own output. If only region i disobeys the order then the war breaks out between i and the center, and the center wins with probability $(1 - x_1 - x_2)/(1 - x_{-i})$, and the winner gets the entire output y_i of region i, leaving zero to the loser. In the event of internal warfare between the two regions, no transfer is made from any region to the center.

Proposition 4 *In this model, given any power allocation that is not fully decentralized, internal warfare occurs at equilibrium and the center receives zero revenue from the regions.*

Proof With the power allocation not fully decentralized, $1 > x_1+x_2$, a region i's probability of defeating the center if it revolts by itself, $x_i/(1-x_{-i})$, is less than its probability of defeating the center if it revolts together with the other region, $x_1 + x_2$. Thus, each region prefers to be united in revolting against the center. Furthermore, in the event of a united revolt, each region i's payoff calculation is the same: it wins with probability $x_1 + x_2$, and if it wins it gets y_i; whereas if it does not revolt then it gets $y_i(1 - s)$. Thus, the center expects no revolt from the regions if and only if $1 - s \geq x_1 + x_2$. It follows that the share s of the output that

---

$-2c_i^3/\left(z_i^3(x_i + \alpha)\right)$, and $D_1 f_i(x_i, \gamma_i(x_i, z_i)) D_2^2 f_i(x_i, \gamma_i(x_i, z_i)) = -c_i^2/\left(2z_i^2(x_i + \alpha)\right)$. Thus (31) and (32) follow, with α playing the role of n. To verify (33), note that, for any $r \in (0, 1)$, the maximum of $\left|D_2^2 f_i(x_i, \gamma_i(x_i, z_i))\right|$, among all $(x_i, z_i) \in [r, 1]^2$, is equal to $2c_i^3/\left(r^3\alpha\right)$, and the minimum equal to $2c_i^3/(1+\alpha)$; hence the ratio between the two is $(1 + \alpha)/\left(r^3\alpha\right)$, which converges to a finite number $1/r^3$ when $\alpha \to \infty$.

the center would optimally choose to demand is equal to

$$s = 1 - x_1 - x_2.$$

Consequently, a region i's gross expected payoff is equal to $x_i\left(f_i(x_i, e_i) + y_{-i}\right)$ if it plans to pillage the other region $-i$, equal to $(x_1 + x_2) f_i(x_i, e_i)$ if it expects peace, and $(1 - x_{-i})\left(f_i(x_i, e_i) + y_{-i}\right)$ if it expects to be pillaged by $-i$. In particular, the gross payoff is equal to the middle branch $(x_1 + x_2) f_i(x_i, e_i)$ only if not pillaging the other region constitutes an equilibrium in the subgame given realized output pair (y_1, y_2):

$$x_1\left(y_1 + y_2\right) \leq (x_1 + x_2) y_1 \quad \text{and} \quad x_2\left(y_1 + y_2\right) \leq (x_1 + x_2) y_2.$$

Sum the two weak inequalities to obtain $(x_1 + x_2)(y_1 + y_2) \leq (x_1 + x_2)(y_1 + y_2)$, which forces both weak inequalities into equalities, i.e., $x_i(y_i + y_{-i}) = (x_1 + x_2) y_i$, or equivalently $x_i\left(f_i(x_i, e_i) + y_{-i}\right) = (x_1 + x_2) f_i(x_i, e_i)$, for each $i \in \{1, 2\}$. Given any y_{-i}, this condition holds for only one value of e_i, denoted $e_i^P(x, y_{-i})$, at which the graphs of the functions $(x_i + x_{-i}) f_i(x_i, \cdot)$ and $x_i\left(f_i(x_i, \cdot) + y_{-i}\right)$ intersect. Thus, the interval of effort levels on which peace prevails collapses into the point $e_i^P(x, y_{-i})$, and region i's gross expected payoff equals

$$f_i^*(x, e_i, y_{-i}) = \begin{cases} x_i\left(f_i(x_i, e_i) + y_{-i}\right) & \text{if } e_i \leq e_i^P(x, y_{-i}) \\ (1 - x_{-i})\left(f_i(x_i, e_i) + y_{-i}\right) & \text{if } e_i \geq e_i^P(x, y_{-i}). \end{cases}$$

Since $x_i < 1 - x_{-i}$, the function $f_i^*(x, \cdot, y_{-i})$ is strictly concave on $[0, e_i^P(x, y_{-i})]$ and on $[e_i^P(x, y_{-i}), \infty)$, with a jump of its derivative at the boundary $e_i^P(x, y_{-i})$. Thus, region i's optimal effort is away from the point $e_i^P(x, y_{-i})$, so the region will either pillage others or be pillaged. With internal warfare occurring for sure, the center gets zero revenue. ■

Hence centralization with such an extortive center ends with a warring equilibrium, which is wasteful by Proposition 2. The implication is that in order for centralization to achieve what it is supposed to achieve, it is necessary to curb the arbitrariness of the center.

7.2 When the Selfish Center Can Commit

The negative observation in the previous subsection is partially due to the assumption that the center does not commit to a tax rate at the outset. With the share of outputs demanded after they are realized, the center would push its share to the limit as far as the centralized power allows. If the center can commit, however, it may want to commit to a less draconian tax rate thereby stimulating some efforts from the regions and reaping greater tax revenues.

Thus consider the following variant of the model: Once the power allocation (x_1, x_2) has been determined, the center chooses a tax rate $s \in [0, 1]$ to commit to, with $s \leq 1 - x_1 - x_2$, such that in the event of peace each region transfers a fraction s of its output to the center, whereas in the event of internal warfare the center receives no transfer.

Consequently, a region i's gross expected payoff, from exerting effort e_i and expecting the other region's output y_{-i}, is equal to $x_i \left(f_i(x_i, e_i) + y_{-i}\right)$ if it plans to pillage $-i$, or $(1 - s)f_i(x_i, e_i)$ if it expects peace, or $(1 - x_{-i}) \left(f_i(x_i, e_i) + y_{-i}\right)$ if it expects to be pillaged.

Lemma 11 *For any $s \in [0, 1 - x_1 - x_2]$, if $x_{-i} > 0 \Rightarrow s_i < x_{-i}$ for each $-i$, then a region i's gross expected payoff is*

$$f_i^*(x, e_i, y_{-i}) = \begin{cases} x_i \left(f_i(x_i, e_i) + y_{-i}\right) & \text{if } e_i \leq e_i^L(x, y_{-i}, s) \\ (1 - s)f_i(x_i, e_i) & \text{if } e_i^L(x, y_{-i}, s) \leq e_i \leq e_i^H(x, y_{-i}, s) \\ (1 - x_{-i}) \left(f_i(x_i, e_i) + y_{-i}\right) & \text{if } e_i \geq e_i^H(x, y_{-i}, s), \end{cases} \tag{34}$$

with the thresholds $e_i^L(x, y_{-i}, s)$ and $e_i^H(x, y_{-i}, s)$ defined analogously to $e_i^L(x, y_{-i})$ and $e_i^H(x, y_{-i})$ in Eqs. (3) and (4) so that $e_i^H(x, y_{-i}, s) = \infty$ if $x_{-i} = 0$.

Proof Appendix F. ■

Receiving zero tax revenue in the event of internal warfare, the center weakly prefers a peaceful equilibrium to any warring one. At any peaceful equilibrium, each region i solves

$$\max_{e_i' \in \mathbb{R}_+} (1 - s)f_i(x_i, e_i') - c_i e_i', \tag{35}$$

which, by Lemma 6, admits a unique solution, denoted by $\phi_i(x_i, s)$, continuous in (x_i, s), as

$$\phi_i(x_i, s) = \gamma_i(x_i, 1 - s). \tag{36}$$

Based on analyses of ϕ_i, the next proposition says that, as long as the power allocation admits a peaceful equilibrium without binding the incentive constraint in the original model, peace is attainable in the current model despite a selfish center. Furthermore, the center's preference of peace over war is strict, and so is its preference of making a commitment.

Proposition 5 *Given any power allocation (x_1, x_2) such that Ineq. (18) holds strictly for each $i \in \{1, 2\}$, there exists $s \in (0, 1 - x_1 - x_2]$ that admits a peaceful equilibrium in the* *current model, and the center strictly prefers committing to such an s to no commitment or any warring equilibrium.*

Proof Appendix F. ■

To reap more tax revenues, the selfish center would commit to a tax rate less draconian than in the previous subsection, thereby encouraging productive efforts from the regions.

Proposition 6 *Given any power allocation (x_1, x_2) such that Ineq. (18) holds strictly for each $i \in \{1, 2\}$, the center's revenue-maximizing tax rate exists and belongs to $(0, 1-x_1-x_2)$.*

Proof Appendix F. ■

This modified model for simplicity gives all bargaining power to the selfish center. Thus one would expect that in general the normative foundation of centralization is ruined by such a greedy, powerful center. Yet the next remark says that there are cases where even the best among decentralized systems is outperformed by even the worst centralized system, the total centralization, where power is entirely taken from the regions by a selfish center, who commits to a tax rate to maximize its own revenues.

Remark 3 *Suppose, for each $i \in \{1, 2\}$ and any $(x_i, e_i) \in [0, 1] \times \mathbb{R}_+$,*

$$f_i(x_i, e_i) = \ln(x_i + 1) + 4\ln(e_i + 1), \tag{37}$$

with an effort cost $c \in (0, 4)$ identical to both regions i. Then:

a. if $c \in (0.36, 3.64)$ then, at the social-surplus maximum among fully decentralized systems, the power allocation either gives approximately 0.09 of the power to one region and the rest to the other region, or gives each region 1/2 of the total power;

b. given the total centralization allocation, $(0, 0)$, the center strictly prefers to commit to a tax rate rather than choose it after the outputs have been produced;

c. there exists a nonempty nondegenerate set of values of c such that the optimal decentralized system in (a) produces less social surplus than the total centralization, with the center committing to a tax rate that maximizes the revenue extracted from the regions.

Proof Appendix G. ■

Centralization is costly, meant to maintain internal peace at the expense of individual productivity. Total centralization, or the Hobbesian Leviathan, goes to the extreme of stripping all power from social members. Such a tremendous cost is compounded by the avarice of the center enjoying absolute power. Yet even such an extreme form of centralization can sometimes outperform even the best among decentralized systems, says the above remark. There, the center prefers to commit than not commit to a tax rate, because without commitment the powerless regions would anticipate that all their outputs will be extorted by the center and hence would exert no effort at all, rendering zero tax revenue for the center. By a similar token, the center would rather commit to a tax rate that leaves the regions an ample fraction of their outputs for them to exert efforts. When the marginal product of efforts is sufficiently larger than that of power in this example, the social loss from depriving the regions of their power is more than covered by the social gain from the efforts that the regions exert in anticipation of internal peace.

Naturally one would ask whether a center's commitment is credible and how it can be made credible. As in Myerson [40], where the prince is assumed to be able to commit to mechanisms that randomize the penalty on the governor, one could interpret the importance of the commitment assumption as an implication that the theory points to the necessity of institutionalized monitoring of the prince there, and the center here.[18] Our model may allow for extensions that incorporate such institutional structures of the center, thereby to produce deeper assessments on the normative foundation of centralization.

A Proofs of Lemmas 4–6 and Formulas (10)–(13)

Lemma 4 In exerting effort e_i region i incurs the sunk cost $c_i e_i$ and produces output $y_i := f_i(x_i, e_i)$. In the endgame given its output y_i and the other's y_{-i}, region i's (gross) expected payoff is calculated below for each case listed on the right-hand side of Eq. (1).

[18] Acemoglu and Robinson [4] regard the ruling elites' lack of commitment ability to transfer their power to the mass as a fundamental problem in undemocratic societies. In our setting, however, the commitment is not about transferring power, but rather committing to a tax rate. There are anecdotes in history where powerful rulers committed to their own policies. Examples include Shang Yang's reform of the Qin Kingdom in China around 350 BCE, where he implemented a strict penal code by applying it first to the nobles, and Augustus's implementation of his marriage laws, which he himself observed through adopting enough children to meet the requirement of the laws and later exiling his own daughter for violation of the laws.

Case 1: $(1-x_i)y_i < x_i y_{-i}$, i.e., $x_i(y_i+y_{-i}) > y_i$, which implies, by Lemma 2, that $x_{-i}(y_{-i}+y_i) < y_{-i}$. Thus, by Lemma 1.b, region $-i$ strictly prefers to not pillage i, hence region i strictly prefers to pillage $-i$. Thus i's expected payoff equals $x_i(y_i+y_{-i})$, and $-i$'s equals $(1-x_i)(y_i+y_{-i})$.

Case 2: $x_{-i}y_i > (1-x_{-i})y_{-i}$, which is symmetric to Case 1, with the roles of i and $-i$ switched. Hence region i's expected payoff equals $(1-x_{-i})(y_i+y_{-i})$.

Case 3: $(1-x_i)y_i > x_i y_{-i}$ and $x_{-i}y_i < (1-x_{-i})y_{-i}$. The first inequality implies that region i strictly prefers to not pillage the other region $-i$ whether $-i$ pillages i or not (Lemma 1.b), and the second inequality implies the same strict preference from the perspective of region $-i$. Hence peace prevails and region i's payoff equals y_i.

Case 4: $(1-x_i)y_i = x_i y_{-i}$, i.e., $x_i(y_i+y_{-i}) = y_i$. Either (i) $x_i + x_{-i} < 1$ or (ii) $x_i + x_{-i} = 1$. In subcase (i), the equality $x_i(y_i+y_{-i}) = y_i$ coupled with Lemma 2 implies $x_{-i}(y_i+y_{-i}) < y_{-i}$, hence region $-i$ strictly prefers to not pillage i (Lemma 1.b); consequently, the expected payoff for region i is equal to $x_i(y_i+y_{-i}) = y_i$, whether it pillages $-i$ or not; and the expected payoff for the other region $-i$ equals y_{-i} or $(1-x_i)(y_i+y_{-i})$, depending on whether i pillages $-i$ or not. In subcase (ii), $x_i = 1-x_{-i} = x_i/(x_i+x_{-i})$. Thus, in all the four cells of the matrix in the proof of Lemma 1, region i's payoffs are equal to one another, and the other region $-i$'s payoffs equal to either $(1-x_i)(y_i+y_{-i})$ or y_{-i}.

Case 5: $x_{-i}y_i = (1-x_{-i})y_{-i}$, which is symmetric to Case 4, with the roles of i and $-i$ switched. Thus, region i's expected payoff is either $(1-x_{-i})(y_i+y_{-i})$ or y_i, which are equal to each other, since $x_{-i}y_i = (1-x_{-i})y_{-i}$ means $y_i = (1-x_{-i})(y_i+y_{-i})$. ■

Lemma 5 Define $e_i^L(x, y_{-i})$ and $e_i^H(x, y_{-i})$ according to Eqs. (3) and (4). With $f_i(x_i,\cdot)$ continuous and strictly decreasing, $(1-x_i)f_i\left(x_i, e_i^L(x,y_{-i})\right) \leq x_i y_{-i}$, which implies, by the fact $x_i \leq 1-x_{-i}$, that $x_{-i}f_i\left(x_i, e_i^L(x,y_{-i})\right) \leq (1-x_{-i})y_{-i}$, hence $e_i^L(x,y_{-i}) \leq e_i^H(x,y_{-i})$. Then Eq. (5) follows from Eq. (1), Claims (a) and (b) follow from the assumption that $f_i(x_i,\cdot)$ is strictly increasing, and (c) from the assumed strict concavity of $f_i(x_i,\cdot)$.

To prove (d), repeat the above proof of $e_i^L(x,y_{-i}) \leq e_i^H(x,y_{-i})$ based on the stronger hypothesis $x_i < 1-x_{-i}$ and we have $x_1 + x_2 < 1 \Rightarrow e_i^L(x,y_{-i}) < e_i^H(x,y_{-i})$. To prove the reverse, suppose $x_1 + x_2 = 1$ and pick any $e_i > e_i^L(x,y_{-i})$. Then Claim (a) implies $x_i(f_i(x_i,e_i)+y_{-i}) < f_i(x_i,e_i)$, which means, with $x_i = 1-x_{-i}$, that $(1-x_{-i})(f_i(x_i,e_i)+y_{-i}) < f_i(x_i,e_i)$. Hence Claim (b) implies $e_i > e_i^H(x,y_{-i})$. Thus $e_i^L(x,y_{-i}) \geq e_i^H(x,y_{-i})$ and hence

$e_i^L(x, y_{-i}) = e_i^H(x, y_{-i})$. Thus $x_1 + x_2 = 1 \Rightarrow e_i^L(x, y_{-i}) = e_i^H(x, y_{-i})$, hence (d) follows.

To prove (e), note, by Claim (a), that $e_i < e_i^L(x, y_{-i})$ is equivalent to $x_i(y_i + y_{-i}) > y_i$, i.e., $(1 - x_i)(y_i + y_{-i}) < y_{-i}$; that is equivalent to the condition that e_{-i} (which produces y_{-i}) belongs to the interval over which the function $f_{-i}(x_{-i}, \cdot)$ is above the function $(1 - x_i)(f_{-i}(x_{-i}, \cdot) + y_i)$, which by Claim (b), with the roles of i and $-i$ switched, is the same as $e_{-i} > e_{-i}^H(x, y_i)$. The proof of $e_i \le e_i^L(x, y_{-i}) \Leftrightarrow e_{-i} \ge e_{-i}^H(x, y_i)$ is likewise, and so is $e_i = e_i^L(x, y_{-i}) \Leftrightarrow e_{-i} = e_{-i}^H(x, y_i)$. ■

Lemma 6 For Claim (a), uniqueness of the solution follows directly from the assumed strict concavity of $f_i(x_i, \cdot)$, and existence follows from the assumption that $f_i(x_i, \cdot)$ is continuously differentiable and $\lim_{e_i \to \infty} D_2 f_i(x_i, e_i) < c_i$. To prove Claim (b), let

$$\zeta_i(x_i) := \inf \left\{ z_i \in [0, 1] : \lim_{e_i \downarrow 0} z_i D_2 f_i(x_i, e_i) > c_i \right\}.$$

Note $\zeta_i(x_i) < 1$ by the assumption $\lim_{e_i \downarrow 0} D_2 f_i(x_i, e_i) > c_i$. Since $D_2 f_i(x_i, \cdot)$ is assumed strictly decreasing, if $z_i \le \zeta_i(x_i)$ then $\gamma_i(x_i, z_i) = 0$, and if $z_i > \zeta_i(x_i)$ then $\gamma_i(x_i, z_i) > 0$ and the first-order condition $z_i D_2 f_i(x_i, \gamma_i(x_i, z_i)) = c_i$ is satisfied. Hence Claim (b) is proved. Next we claim that $\gamma_i(x_i, z_i)$ is weakly increasing in (x_i, z_i). The reason is that the objective function of the maximization problem in Eq. (6) exhibits increasing difference in (e_i, x_i), because $D_1 D_2 f_i \ge 0$ by assumption, and increasing difference in (e_i, z_i), because $D_2 f_i > 0$ by assumption. Thus, to prove Claim (c) it suffices to prove that $\gamma_i(x_i, \cdot)$ is strictly increasing on $(\zeta_i(x_i), 1]$, on which $\gamma_i(x_i, \cdot)$ is determined by the first-order condition. This condition implies, with $D_2 f_i(x_i, \cdot)$ assumed strictly decreasing, that a higher z_i within this interval means a higher $\gamma_i(x_i, z_i)$. Hence strict monotonicity on this interval is proved, and so is Claim (c). To prove Claim (d), note from the monotonicity of γ_i that $\gamma_i(x_i, z_i) \le \gamma_i(1, 1)$ for any $(x_i, z_i) \in [0, 1]^2$, hence the choice set $\mathbb{R}_+$ in Eq. (6) can be replaced by the compact set $[0, \gamma_i(1, 1)]$. The theorem of maximum therefore applies to the maximization problem in Eq. (6) and implies the continuity of γ_i, proving Claim (d). ■

Proof of (10), (11), (12) and (13) For each $x_i \in [0, 1)$, Ineq. (10) holds because $\gamma_i(x_i, \cdot)$ is weakly increasing on $[0, 1]$ and strictly increasing on a neighborhood of one (Lemma 6.c). By Lemma 7, $\psi_i(x_i)$ is the unique solution of the problem in (7), Thus, (10) implies (11). Eq. (12) follows from Eq. (9), the implicit function theorem and the assumption $D_2^2 f_i < 0$.

Apply the Milgrom-Segal envelope theorem [33] to the problem in Eq. (7) to obtain the equality in (13), where the inequality is due to $D_1 f_i > 0$ by assumption. ■

B Proofs of Theorem 2, Corollary 3 and Remark 1

Theorem 2 To prove existence of such a maximum, we apply the Weierstrass theorem to the problem (20). There, the objective function is continuous because f_i is continuous by assumption and ψ_i continuous by Lemma 7. The constraints of the problem define a compact space because those (x_1, x_2) satisfying (18) constitute a closed subset of the compact space $\{(x_1, x_2) \in [0, 1]^2 : x_1 + x_2 \leq 1\}$, and that subset is closed because f_i and ψ_i, as well as η_i by Lemma 7, are each continuous. In addition, the compact space defined by the constraints is nonempty because it contains $(0, 0)$ (Corollary 1). Thus, a solution for problem (20) exists.

For the rest of the theorem, note that the objective in Problem (20) is strictly increasing in (x_1, x_2) by the inequality in (13). Also note, from the "if" part of Proposition 3, that the constraint $x_1 + x_2 \leq 1$ is not binding at the maximum. Thus, if the constraint (18) is non-binding for both regions $i \in \{1, 2\}$, with f_j, ψ_j and η_j continuous for each j, a slight increase of both coordinates yields larger surplus than the maximum, a contradiction. Thus, for some region i, the constraint (18) is binding, i.e., Eq. (21) holds. If the constraint is binding for both regions then we have Case (i) in the theorem. Else, say the constraint is binding only for i. Consequently, with the objective strictly increasing and $x_1 + x_2 < 1$, the maximum is a tangent point between the curve defined by the binding constraint (18) for i and an indifference curve of the objective function, i.e., we have Case (ii) of the theorem, unless the maximum is not interior to $\mathbb{R}^2_+$, i.e., $x_{-i} = 0$, Case (iii) of the theorem. ■

Corollary 3 Note that $\rho_i > 0$ by the definition of ρ_i (Eq. (22)) and the proof of Lemma 9. By Theorem 2, at any optimal centralized power allocation (x_1, x_2), Eq. (21) holds for some $i \in \{1, 2\}$. Thus, it suffices to prove for any $i \in \{1, 2\}$

$$(x_i, x_{-i}) \text{ satisfies Eq. (21)} \Rightarrow x_i \geq \rho_i. \tag{38}$$

To prove (38), note that Eq. (21) implies

$$\begin{aligned} f_i(x_i, \psi_i(x_i)) - c_i\psi_i(x_i) &\leq x_i \left(f_i(x_i, \eta_i(x_i)) + f_{-i}(x_{-i}, \psi_{-i}(x_{-i}))\right) \\ &\overset{(10)}{\leq} x_i \left(f_i(x_i, \psi_i(x_i)) + f_{-i}(x_{-i}, \psi_{-i}(x_{-i}))\right). \end{aligned}$$

Hence

$$x_i \geq \frac{f_i(x_i,\psi_i(x_i)) - c_i\psi_i(x_i)}{f_i(x_i,\psi_i(x_i)) + f_{-i}(x_{-i},\psi_{-i}(x_{-i}))} > \frac{f_i(0,\psi_i(0)) - c_i\psi_i(0)}{f_i(1,\psi_i(1) + f_{-i}(1,\psi_{-i}(1))},$$

where the last inequality follows from the monotonicity of $f_i(\cdot,\psi_i(\cdot)) - c_i\psi_i(\cdot)$ and $f_i(\cdot,\psi_i(\cdot))$ due to (12) and (13). Hence (38) is proved, as desired. ■

Remark 1 Consider the case where, for some $\alpha > 0$, the production function is given by Eq. (26) for each region i. Then for each $i \in \{1,2\}$ and for any $(x_i, e_i) \in [0,1] \times \mathbb{R}_+$

$$\frac{D_1 f_i(x_i,e_i)}{-D_2 f_i(x_i,e_i)} = -\frac{(1/2)\sqrt{e_i/(x_i+\alpha)}}{(1/2)\sqrt{(x_i+\alpha)/e_i}} = -\frac{e_i}{x_i+\alpha},$$

which goes to zero as $\alpha \to \infty$. Whereas, by Eq. (22),

$$\rho_i = \frac{f_i(0,\psi_i(0)) - c_i\psi_(0)}{f_1(1,\psi_1(1)) + f_2(1,\psi_2(1))} \overset{(41)}{=} \frac{\alpha/(2c_i) - \alpha/(4c_i)}{(1+\alpha)/2c_1 + (1+\alpha)/(2c_2)} = \frac{\alpha}{1+\alpha}\cdot\frac{c_{-i}}{2(c_1+c_2)},$$

which converges to the positive constant $c_{-i}/(2(c_1+c_2))$ as $\alpha \to \infty$. ■

C Proof of Theorem 3

By the definition of symmetric allocations and Lemma 10, $\tilde{x}(t)$ is a solution of

$$\begin{array}{ll} \max\limits_{x\in[0,1/2]} & f\left(x,\psi(x,t);t\right) - c\psi(x,t) \\ \text{subject to} & \sigma(x;t) \geq 0, \end{array} \tag{39}$$

where

$$\sigma(x;t) := f\left(x,\psi(x,t);t\right) - c\psi(x,t) - x\left(f\left(x,\eta(x,t);t\right) + f\left(x,\psi(x,t);t\right)\right) + c\eta(x,t).$$

The envelope theorem, applied to the problems in (7) and (8) with parameter t, gives

$$\begin{aligned} \frac{\partial}{\partial t}\sigma(x;t) &= D_3 f\left(x,\psi(x,t);t\right) - xD_3 f\left(x,\eta(x,t);t\right) \\ &\quad -x\left(D_3 f\left(x,\psi(x,t);t\right) + D_2 f\left(x,\psi(x,t);t\right)\frac{\partial}{\partial t}\psi(x,t)\right) \\ &\overset{(9)}{=} (1-x)D_3 f\left(x,\psi(x,t);t\right) - x\left(D_3 f\left(x,\eta(x,t);t\right) + c\cdot\frac{\partial}{\partial t}\psi(x,t)\right) \\ &= (1-2x)D_3 f\left(x,\eta(x,t);t\right) \\ &\quad +(1-x)\left[D_3 f\left(x,\psi(x,t);t\right) - D_3 f\left(x,\eta(x,t);t\right)\right] - xc\cdot\frac{\partial}{\partial t}\psi(x,t) \\ &= (1-2x)D_3 f\left(x,\eta(x,t);t\right) \\ &\quad +(1-x)D_2 D_3 f\left(x,\xi;t\right)\left(\psi(x,t) - \eta(x,t)\right) - xc\cdot\frac{\partial}{\partial t}\psi(x,t) \end{aligned}$$

for some ξ between $\psi(x,t)$ and $\eta(x,t)$. Analogously to Eq. (12), we have

$$\frac{\partial}{\partial t}\psi(x,t) = \frac{D_3 D_2 f\left(x, \psi(x,t); t\right)}{-D_2^2 f\left(x, \psi(x,t); t\right)}. \tag{40}$$

Thus, by the fact $\psi(x,t) > \eta(x,t)$ (Ineq. (10)), Ineq. (25) and $0 \le x \le 1/2$,

$$\begin{aligned} \frac{\partial}{\partial t}\sigma(x;t) &\geq (1-2x) D_3 f\left(x, \eta(x,t); t\right) + \frac{1}{2}\|D_2 D_3 f\|_{\inf}\|\psi - \eta\|_{\inf} - \frac{1}{2}\frac{c\|D_3 D_2 f\|_{\sup}}{\|-D_2^2 f\|_{\inf}} \\ &\geq (1-2x) D_3 f\left(x, \eta(x,t); t\right), \end{aligned}$$

which is nonnegative due to the fact $1 - 2x \geq 0$ and assumption $D_3 f > 0$. Furthermore, the last inequality is strict for any $x \in [0, 1/2]$ such that $\sigma(x;t) \geq 0$, because for any such x the symmetric allocation (x,x) satisfies Ineq. (18) for both i and hence $2x < 1$ (Proposition 3). Thus, the choice set $\{x \in [0,1/2] : \sigma(x;t) \geq 0\}$ is strictly increasing (in strong-set order) in t.

The objective in (39) obeys the single crossing difference condition in (x,t), because

$$\begin{aligned} \frac{\partial^2}{\partial t \partial x}\left(f\left(x, \psi(x,t); t\right) - c\psi(x,t)\right) &\overset{(13)}{=} \frac{d}{dt} D_1 f\left(x, \psi(x,t); t\right) \\ &= D_2 D_1 f\left(x, \psi(x,t); t\right) \frac{\partial}{\partial t}\psi(x,t) + D_3 D_1 f\left(x, \psi(x,t); t\right), \end{aligned}$$

which is nonnegative because both $D_2 D_1 f$ and $D_3 D_1 f$ are nonnegative by assumption and $\frac{\partial}{\partial t}\psi(x,t) \geq 0$ by Eq. (40) and the assumptions $D_3 D_2 f \geq 0$ and $D_2^2 f \leq 0$.

Thus, by the Milgrom-Shannon theorem [34], the set of solutions for the problem (39) is increasing in t. The solution for that problem is unique, because the objective is strictly increasing in the choice variable x (Eq. (13) and the assumption $D_1 f > 0$). Hence $t'' > t'$ implies $\tilde{x}(t'') \geq \tilde{x}(t')$. Furthermore, $\tilde{x}(t'') \neq \tilde{x}(t')$. That is because the constraint $\sigma(x,t) \geq 0$ is not binding when $x = \tilde{x}(t')$ and $t = t''$, as $\sigma\left(\tilde{x}(t'), t''\right) > \sigma\left(\tilde{x}(t'), t'\right) \geq 0$ given the fact that $\sigma(x, \cdot)$ is strictly increasing. Thus, the objective in (39) given $t = t''$, a strictly increasing function, is larger at some feasible $x \neq \tilde{x}(t')$ than at $\tilde{x}(t')$. That completes the proof.

D Proof of Remark 2

Claim (a) By Eq. (26), $D_2 f_i(x_i, e_i) = \frac{1}{2}\sqrt{(x_i+\alpha)/e_i}$. Hence the solutions ψ_i and η_i for the problems in (7) and (8) are identical for both i, so we suppress their subscript i:

$$\begin{aligned} \psi(x_i) &= (x_i+\alpha)/(4c^2), \\ \eta(x_i) &= (x_i+\alpha)x_i^2/(4c^2); \\ f_i(x_i,\psi(x_i)) &= (x_i+\alpha)/(2c), \\ f_i(x_i,\eta(x_i)) &= (x_i+\alpha)x_i/(2c). \end{aligned} \tag{41}$$

By Theorem 1, the optimal decentralized system maximizes over $x_1 \in [0,1]$ the social surplus

$$S_{\mathrm{dcn}}(x_1) = \frac{1}{4c}\left((x_1+\alpha)(2x_1 - x_1^2) + (1+\alpha-x_1)(1-x_1^2)\right).$$

Hence

$$4c\frac{d}{dx_1}S_{\mathrm{dcn}}(x_1) = (2-4\alpha)x_1 + 2\alpha - 1,$$

which is strictly decreasing in x_1 if $\alpha > 1/2$, and strictly increasing in x_1 if $\alpha < 1/2$. When $\alpha < 1/2$, S_{dcn} is strictly convex and the two endpoints, $x_1 = 0$ and $x_1 = 1$, are the maximums. When $\alpha > 1/2$, S_{dcn} is strictly concave and the maximum is the solution of $\frac{d}{dx_1}S_{\mathrm{dcn}}(x_1) = 0$, which is $x_1 = 1/2$. Thus the maximal surplus generated by fully decentralized systems is

$$S^*_{\mathrm{dcn}} = \begin{cases} \frac{1+\alpha}{4c} & \text{if } \alpha \le 1/2 \\ \frac{3}{8c}\left(\frac{1}{2}+\alpha\right) & \text{if } \alpha \ge 1/2. \end{cases} \tag{42}$$

By the above calculation of $\psi(x_i)$ and Eq. (20), the social surplus generated by a peaceful system that is not fully decentralized is equal to

$$S_{\mathrm{cnt}}(x_1,x_2) := \sum_{i=}^{2}\left(\frac{x_i+\alpha}{2c} - \frac{x_i+\alpha}{4c}\right) = \frac{1}{4c}\left(x_1+x_2+2\alpha\right), \tag{43}$$

and Ineq. (18), the incentive constraint for peace, becomes

$$(x_i+\alpha)(1-x_i^2) \ge 2x_i(x_{-i}+\alpha)$$

for region i. Comparing (42) with (43) we see that, when $\alpha = 0$, $S^*_{\mathrm{dcn}} \ge S_{\mathrm{cnt}}(x_1,x_2)$ for any power allocation (x_1,x_2), and the inequality is strict unless $x_1 + x_2 = 1$, i.e., unless (x_1,x_2) is fully decentralized. Thus, when $\alpha = 0$, the optimal decentralized system generates higher social surplus than any centralized system.

To prove the rest of the claim, we compare the optimal decentralized system with a centralized one with power allocation $x_1 = x_2 = \sqrt{2}-1$. It is easy to verify that this allocation satisfies the above incentive constraint for peace. The allocation generates a social surplus

$$S^o := S_{\text{cnt}}\left(\sqrt{2}-1, \sqrt{2}-1\right) = \frac{\sqrt{2}-1+\alpha}{2c}.$$

This equation, combined with Eq. (42), implies

$$\begin{aligned}
\alpha \leq 1/2 \quad &\Rightarrow \quad \left[S^*_{\text{dcn}} > (\geq)\, S^o \Leftrightarrow \alpha < (\leq)\, 3-2\sqrt{2}\right], \\
\alpha \geq 1/2 \quad &\Rightarrow \quad \left[S^o > S^*_{\text{dcn}} \Leftrightarrow \alpha > 11/2-4\sqrt{2}\right].
\end{aligned}$$

Thus, $S^*_{\text{dcn}} > S^o$ if $\alpha < 3-2\sqrt{2}$, and $S^*_{\text{dcn}} < S^o$ if $\alpha > 3-2\sqrt{2}$. That proves the claim.

Claim (b) By Eq. (27), $D_2 f_i(x_i, e_i) = (\theta x_i)^\beta \tau e_i^{\tau-1}$. Hence

$$\begin{aligned}
\psi(x_i) &= \left(\frac{\tau}{c}(\theta x_i)^\beta\right)^{1/(1-\tau)}, \\
\eta(x_i) &= \left(\frac{\tau}{c}\theta^\beta x_i^{1+\beta}\right)^{1/(1-\tau)}; \\
f_i\left(x_i, \psi(x_i)\right) &= \left(\frac{\tau}{c}\right)^{\tau/(1-\tau)} (\theta x_i)^{\beta/(1-\tau)}, \\
f_i\left(x_i, \eta(x_i)\right) &= \left(\frac{\tau}{c}\right)^{\tau/(1-\tau)} \theta^{\beta/(1-\tau)} x_i^{(\beta+\tau)/(1-\tau)}.
\end{aligned}$$

As in Claim (a), the social surpluses from a fully decentralized system, and a centralized one, are respectively equal to

$$S_{\text{dcn}}(x_1, x_2) = \theta^{\beta/(1-\tau)}\left(\frac{\tau}{c}\right)^{\tau/(1-\tau)} \sum_{i=1}^{2}(1-\tau x_i)x_i^{(\beta+\tau)/(1-\tau)}, \tag{44}$$

$$S_{\text{cnt}}(x_1, x_2) = \theta^{\beta/(1-\tau)}\left(\frac{\tau}{c}\right)^{\tau/(1-\tau)} (1-\tau) \sum_{i=1}^{2} x_i^{\beta/(1-\tau)}. \tag{45}$$

Claim (b1): If $\beta + \tau \geq 1$ then $S_{\text{dcn}}(1, 0) > S_{\text{cnt}}(x_1, x_2)$ for any allocation (x_1, x_2) that is not fully decentralized. To prove that, note from $\beta + \tau \geq 1$ that $\beta/(1-\tau) \geq 1$. This, coupled with the fact that $x_i^{\beta/(1-\tau)}$ is a strictly increasing function of x_i, implies

$$x_1^{\beta/(1-\tau)} + x_2^{\beta/(1-\tau)} < x_1^{\beta/(1-\tau)} + (1-x_1)^{\beta/(1-\tau)} \leq 1.$$

Thus, with $\theta > 0$ and $0 < \tau < 1$ assumption, Claim (b1) follows from Eqs. (44)–(45).

Claim (b2): The allocation $(1,0)$ is social-surplus maximizing among all fully decentralized allocations. To prove that, let $n := 1/(1-\tau)$. Then one can show that

$$(1-\tau x_i)x_i^{(\beta+\tau)/(1-\tau)} = (1-(1-1/n)x_i)\, x_i^{n-1+n\beta} \leq 1/n$$

for any $x_i \in [0,1]$, with the weak inequality being equality if and only if $x_i = 1$.[19] Hence Claim (b2) follows from Eq. (44).

Claim (b3): The power allocation $(1/(2n), 1/(2n))$, with $n := 1/(1-\tau)$, satisfies the incentive constraint for peace, Ineq. (18) for each i. To prove that, note from the assumptions $\theta > 0$ and $\tau > 0$ that Ineq. (18) is equivalent to

$$(1-\tau)x_i^{\beta/(1-\tau)}\left(1 - x_i^{1/(1-\tau)}\right) \geq x_i x_{-i}^{\beta/(1-\tau)}.$$

The claim then follows from the notation $n = 1/(1-\tau)$.

Claim (b4): If $\beta/(1-\tau)$ is sufficiently small, then the fully decentralized allocation $(1,0)$ generates less social surplus than the allocation $(1/(2n), 1/(2n))$, with $n := 1/(1-\tau)$. To prove that, denote $m := 1/(n\beta)$. By Eq. (45),

$$S_{\text{cnt}}\left(\frac{1}{2n}, \frac{1}{2n}\right) = \theta^{\beta/(1-\tau)}\left(\frac{\tau}{c}\right)^{\tau/(1-\tau)}(1-\tau)2\left(\frac{1}{2n}\right)^{\beta/(1-\tau)} = \theta^{\beta/(1-\tau)}\left(\frac{\tau}{c}\right)^{\tau/(1-\tau)}\frac{2}{n}\left(\frac{1}{2n}\right)^{1/m},$$

where $\left(\frac{1}{2n}\right)^{1/m}$ converges to one when $m \to \infty$, i.e., when $\beta/(1-\tau) \to 0$. Whereas, as calculated in Claim (b2), $S_{\text{dcn}}(1,0) = \theta^{\beta/(1-\tau)}\left(\frac{\tau}{c}\right)^{\tau/(1-\tau)}\frac{1}{n}$. Hence $S_{\text{dcn}}(1,0) < S_{\text{cnt}}(1/(2n), 1/(2n))$.

Claims (b1), (b2), (b3) and (b4) combined, Claim (b) is proved.

E Proof of Theorem 4

Lemma 12 *For any environment (f_1, f_2) that satisfies the assumptions in Section 2 there exists a power allocation (x_1, x_2) that satisfies Eq. (21) for each $i \in \{1,2\}$.*

[19] We need only to prove, for all $x_i \in (0,1]$, that $(1-(1-1/n)x_i)\, x_i^{n-1+n\beta} \leq 1/n$, which is equivalent to

$$1-(1-1/n)x_i \leq (1/n)x_i^{-n+1-n\beta}.$$

At $x_i = 1$, the two sides are equal to each other. As x_i increases from zero to one, the left-hand side decreases at a constant rate $1-1/n$, while the right-hand side decreases at a rate $(1-1/n+\beta)x^{-n-n\beta}$ and this rate keeps diminishing down to $1-1/n+\beta$ at $x_i = 1$, which is still larger than that of the left-hand. Hence the inequality is proved.

Proof For any power allocation (x'_1, x'_2) and for each $i \in \{1,2\}$, let $x''_i \in [0,1]$ be a solution for x_i in Eq. (21) where $x_{-i} = x'_{-i}$. Such x''_i exists in $[0,1]$ by the intermediate-value theorem: when $x_i = 0$, the left-hand side of Eq. (21) is strictly positive and the right-hand side is zero (the proof of Lemma 9); when $x_i = 1$, the right-hand side of Eq. (21) is equal to its left-hand side plus the nonnegative term $f_{-i}(x_{-i}, \psi_{-i}(x_{-i}))$, since $\eta_i(1) = \psi_i(1)$; thus, since f_i and f_{-i} are continuous by assumption, and ψ_i, ψ_{-i} and η_i continuous by Lemma 7, a solution x''_i for Eq. (21) exists. Now that (x''_i, x'_{-i}) solves Eq. (21) by playing the role of (x_i, x_{-i}), it follows from the continuity of the functions f_i, f_{-i}, ψ_i, ψ_{-i} and η_i in this equation that x''_i is continuous in x'_{-i}. Thus, we have constructed a continuous mapping $(x'_1, x'_2) \mapsto (x''_1, x''_2)$ from $[0,1]$ to $[0,1]$. By Brouwer's fixed point theorem, the mapping has a fixed point, say (x_1, x_2), which satisfies Eq. (21) for both i and $-i$. ■

Proof of Theorem 4 Pick any sequence $(f_1^n, f_2^n)_{n=1}^{\infty}$ specified by the hypothesis of the theorem. For each n, by Lemma 12, there exists a power allocation, denoted (x_1^n, x_2^n), that satisfies Eq. (21) for each $i \in \{1,2\}$. Then Proposition 3 implies that (x_1^n, x_2^n) constitutes a peaceful system that is not fully decentralized.

For any n in the sequence $(f_1^n, f_2^n)_{n=1}^{\infty}$, denote ψ_i^n for the corresponding ψ_i defined by Eq. (7), η_i^n by Eq. (8), and γ_i^n by Eq. (6). By (6) and the Inada condition $\lim_{e_i \downarrow 0} D_2 f_i^n(x_i, e_i) = \infty$, for any $(x_i, z_i) \in (0,1]^2$ we have $\gamma_i^n(x_i, z_i) > 0$ and

$$D_2 f_i^n(x_i, \gamma_i^n(x_i, z_i)) = c_i / z_i. \tag{46}$$

Thus, the implicit function theorem implies

$$\begin{aligned}
\frac{\partial}{\partial x_i}\gamma_i^n(x_i, z_i) &= \frac{D_1 D_2 f_i^n(x_i, \gamma_i^n(x_i, z_i))}{-D_2^2 f_i^n(x_i, \gamma_i^n(x_i, z_i))}, &(47)\\
\frac{\partial}{\partial z_i}\gamma_i^n(x_i, z_i) &= \frac{D_2 f_i^n(x_i, \gamma_i^n(x_i, z_i))}{-z_i D_2^2 f_i^n(x_i, \gamma_i^n(x_i, z_i))}. &(48)
\end{aligned}$$

For each $i \in \{1,2\}$, with the strictly positive constant $\bar{r}_i$ given by hypothesis (30), let

$$\bar{x}_i := \min\{\bar{r}_i, 1/2\}.$$

By the choice of (x_1^n, x_2^n), the fact (38) and the hypothesis (30), $\lim_n x_i^n \geq \liminf_n \rho_i^n > \bar{r}_i$ for each $i \in \{1,2\}$. Hence there exists N such that, for each $i \in \{1,2\}$, $x_i^n > \bar{r}_i \geq \bar{x}_i$ if $n \geq N$. For any such n, with $f_i^n(\cdot, \psi_i^n(\cdot)) - c_i \psi_i^n(\cdot)$ weakly increasing due to (13),

$$f_i^n(x_i^n, \psi_i^n(x_i^n)) - c_i \psi_i^n(x_i^n) \geq f_i^n(\bar{x}_i, \psi_i^n(\bar{x}_i)) - c_i \psi_i^n(\bar{x}_i) =: S_i^n(\bar{x}_i).$$

Thus it suffices to prove that, for all large n and any fully decentralized allocation (x_1^*, x_2^*),

$$\sum_{i=1}^{2} S_i^n(\bar{x}_i) > S_{\mathrm{dcn}}^n(x_1^*, x_2^*), \tag{49}$$

with $S_{\mathrm{dcn}}^n(x_1^*, x_2^*)$ denoting the social surplus generated by (x_1^*, x_2^*).

To this end, pick any $\delta \in (0, 1/2)$. Since both f_i^n and γ_i^n are differentiable, by the intermediate-value theorem there exists $z_i^n \in (1/2, 1-\delta)$ such that

$$\begin{aligned}
& f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, 1-\delta)) - c_i\gamma_i^n(\bar{x}_i, 1-\delta) - f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, 1/2)) + c_i\gamma_i^n(\bar{x}_i, 1/2) \\
\overset{(48)}{=} \; & \left(\frac{1}{2} - \delta\right)(D_2 f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, z_i^n)) - c_i)\frac{D_2 f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, z_i^n))}{-z_i^n D_2^2 f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, z_i^n))} \\
\overset{(46)}{=} \; & \frac{1-2\delta}{2}\left(\frac{c_i}{z_i^n} - c_i\right)\frac{c_i/z_i^n}{-z_i^n D_2^2 f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, z_i^n))} \\
> \; & \frac{c_i^2\delta(1-2\delta)}{2(1-\delta)^3}\left(-D_2^2 f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, z_i^n))\right)^{-1},
\end{aligned}$$

with the last line due to $z_i^n < 1-\delta$. By the fact that $\frac{c_i^2\delta(1-2\delta)}{2(1-\delta)^3} > 0$ and $(\bar{x}, z_i^n) \in [\min\{\bar{r}_i, 1/2\}, 1]^2$, as well as Ineq. (33) and the assumption $D_2^2 f_i < 0$, there exist $\epsilon_i > 0$ and $N' \geq N$ such that for any $n \geq N'$

$$\frac{c_i^2\delta(1-2\delta)}{2(1-\delta)^3}\left(-D_2^2 f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, z_i^n))\right)^{-1} > \frac{\epsilon_i}{b_i\|D_2^2 f_i^n\|_{\min}},$$

where

$$\|D_2^2 f_i^n\|_{\min} := \min_{(x_i, z_i) \in [\min\{\bar{r}_i, 1/2\}, 1]^2}\left|D_2^2 f_i^n(x_i, \gamma_i^n(x_i, z_i))\right|.$$

This being true for each $i \in \{1, 2\}$, there exists $\epsilon > 0$ such that for $n \geq N'$ and each i

$$f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, 1-\delta)) - c_i\gamma_i^n(\bar{x}_i, 1-\delta) - f_i^n(\bar{x}_i, \gamma_i^n(\bar{x}_i, 1/2)) + c_i\gamma_i^n(\bar{x}_i, 1/2) > \frac{\epsilon}{\|D_2^2 f_i^n\|_{\min}}. \tag{50}$$

Consider any fully decentralized allocation (x_1^*, x_2^*). For each $i \in \{1, 2\}$, Eq. (29) says $\Delta_i(\bar{x}_i, x_i^*) = \Delta_i^s(\bar{x}_i, x_i^*) + \Delta_i^p(\bar{x}_i, x_i^*)$. We claim that, for all sufficiently large n,

$$\Delta_i^p(\bar{x}_i, x_i^*) \geq -(x_i^* - \bar{x}_i)^+ \frac{\epsilon}{3\|D_2^2 f_i^n\|_{\min}} \tag{51}$$

for each $i \in \{1, 2\}$. By definition of $\Delta_i^p(\bar{x}_i, x_i^*)$ at Eq. (29) and the fact $\gamma_i(x_i, 0) = 0$ (Lemma 6.b), $\Delta_i^p(\bar{x}_i, x_i^*) \geq 0$ if $x_i^* = 0$. Thus assume, without loss of generality, that $x_i^* > 0$. Then by Eq. (47), $\Delta_i^p(\bar{x}_i, x_i^*)$ is equal to

$$\begin{aligned}
& -(x_i^* - \bar{x}_i)\left(D_1 f_i^n(x_i, \gamma_i^n(x_i, x_i^*)) + (D_2 f_i^n(x_i, \gamma_i^n(x_i, x_i^*)) - c_i)\frac{D_1 D_2 f_i^n(x_i, \gamma_i^n(x_i, x_i^*))}{-D_2^2 f_i^n(x_i, \gamma_i^n(x_i, x_i^*))}\right) \\
\overset{(46)}{=} \; & -(x_i^* - \bar{x}_i)\left(D_1 f_i^n(x_i, \gamma_i^n(x_i, x_i^*)) + (c_i/x_i^* - c_i)\frac{D_1 D_2 f_i^n(x_i, \gamma_i^n(x_i, x_i^*))}{-D_2^2 f_i^n(x_i, \gamma_i^n(x_i, x_i^*))}\right)
\end{aligned} \tag{52}$$

for some x_i between $\bar{x}_i$ and x_i^*. With $x_i^* \in [0,1]$, the second factor in (52) is nonnegative. Thus, if $x_i^* \leq \bar{x}_i$ then (52) is nonnegative and (51) is true. Consider the case $x_i^* > \bar{x}_i$. Then $x_i^* \geq x_i \geq \bar{x}_i$ and hence $(x_i, x_i^*) \in [\min\{\bar{r}_i, 1/2\}, 1]^2$. Thus,

$$-D_2^2 f_i^n\left(x_i, \gamma_i^n(x_i, x_i^*)\right) \geq \|D_2^2 f_i^n\|_{\min}.$$

This, coupled with the fact

$$\lim_{n\to\infty}\left[D_1 f_i^n\left(x_i, \gamma_i^n(x_i, x_i^*)\right)\left(-D_2^2 f_i^n\left(x_i, \gamma_i^n(x_i, x_i^*)\right)\right) + \left(c_i/x_i^* - c_i\right) D_1 D_2 f_i^n\left(x_i, \gamma_i^n(x_i, x_i^*)\right)\right] = 0$$

due to Eqs. (31–32) and $x_i^* > \bar{x}_i \geq \bar{r}_i$, implies that the second factor of the expression (52) is less than $\epsilon/\left(3\|D_2^2 f_i^n\|_{\min}\right)$ for all sufficiently large n. Thus, there exists $N'' \geq N'$ such that Ineq. (51) is true for all $n \geq N''$.

Thus pick any $n \geq N''$. Given the environment (f_1^n, f_2^n), for any fully decentralized power allocation (x_1^*, x_2^*), Ineq. (51) is satisfied and

$$\Delta_1^p(\bar{x}_1, x_1^*) + \Delta_2^p(\bar{x}_2, x_2^*) > -\left((x_1^* - \bar{x}_1)^+ + (x_2^* - \bar{x}_2)^+\right)\frac{\epsilon}{3\|D_2^2 f_i^n\|_{\min}} > -\frac{\epsilon}{\|D_2^2 f_i^n\|_{\min}}.$$

Since (x_1^*, x_2^*) is a power allocation, $x_i^* \leq 1/2$ for some $i \in \{1,2\}$. For such i,

$$\begin{aligned}
\Delta_i^s(\bar{x}_i, x_i^*) &= f_i^n\left(\bar{x}_i, \gamma_i^n(\bar{x}_i, 1)\right) - c_i\gamma_i^n(\bar{x}_i, 1) - f_i^n\left(\bar{x}_i, \gamma_i^n(\bar{x}_i, x_i^*)\right) + c_i\gamma_i^n(\bar{x}_i, x_i^*) \\
&\geq f_i^n\left(\bar{x}_i, \gamma_i^n(\bar{x}_i, 1-\delta)\right) - c_i\gamma_i^n(\bar{x}_i, 1-\delta) - f_i^n\left(\bar{x}_i, \gamma_i^n(\bar{x}_i, 1/2)\right) + c_i\gamma_i^n(\bar{x}_i, 1/2) \\
&\overset{(50)}{>} \frac{\epsilon}{\|D_2^2 f_i^n\|_{\min}},
\end{aligned}$$

with the second line due to the fact that $f_i^n(x_i, \gamma_i^n(x_i, z_i)) - c_i\gamma_i^n(x_i, z_i)$ is weakly increasing in z_i.[20] This monotonicity property also implies $\Delta_{-i}^s(\bar{x}_{-i}, x_{-i}^*) \geq 0$. Thus,

$$\begin{aligned}
\Delta_1(\bar{x}_1, x_1^*) + \Delta_2(\bar{x}_2, x_2^*) &\geq \Delta_i^s(\bar{x}_i, x_i^*) + \Delta_1^p(\bar{x}_1, x_1^*) + \Delta_2^p(\bar{x}_2, x_2^*) \\
&> \frac{\epsilon}{\|D_2^2 f_i^n\|_{\min}} - \frac{\epsilon}{\|D_2^2 f_i^n\|_{\min}} = 0.
\end{aligned}$$

This being true for all fully decentralized allocations (x_1^*, x_2^*), Ineq. (49) is proved, as desired.

F Proofs of Lemma 11 and Propositions 5 and 6

Lemma 11 Since $f_i(x_i, \cdot)$ is strictly increasing and concave, and $y_{-i} \geq 0$, it suffices to have $x_i < 1 - s$ and $1 - s > 1 - x_{-i}$. The former, $x_i < 1 - s$, is part of the hypothesis of

[20] To prove this monotonicity property, differentiate $f_i^n(x_i, \gamma_i^n(x_i, z_i)) - c_i\gamma_i^n(x_i, z_i)$ with respect to z_i, then apply Eqs. (46) and (48) and the assumption $D_2^2 f_i \leq 0$.

the lemma; the latter is implied by the hypothesis because if $x_{-i} = 0$ then according to the statement of the lemma we do not need to prove the bottom branch of Eq. (34). ■

Proposition 5 For each $i \in \{1,2\}$ define

$$\bar{s}_i := \sup\left\{ s \in [0,1] : \lim_{e_i \downarrow 0} (1-s) D_2 f_i(x_i, e_i) \geq c_i \right\}.$$

By Eq. (36) and Lemma 6.b, $\bar{s}_i = 1 - \zeta_i(x_i)$ and $\bar{s}_i > 0$; furthermore, if $s \geq \bar{s}_i$ then $\phi_i(x_i, s) = 0$, and if $s < \bar{s}_i$ then $\phi_i(x_i, s) > 0$ and

$$(1-s) D_2 f_i\left(x_i, \phi_i(x_i, s)\right) = c_i.$$

As long as Eq. (34) is satisfied, one can mimick the proof of Proposition 3 to obtain a sufficient condition for $(x_1, x_2; s)$ to admit a peaceful equilibrium in the current model:

$$(1-s) f_i\left(x_i, \phi_i(x_i,s)\right) - c_i \phi_i(x_i,s) \geq x_i\left(f_i\left(x_i, \eta_i(x_i)\right) + f_{-i}\left(x_{-i}, \phi_{-i}(x_{-i},s)\right)\right) - c_i \eta_i(x_i) \quad (53)$$

for each $i \in \{1,2\}$. We claim that there exists a $\delta \in (0, 1-x_1-x_2]$ such that, for each region i and any $s \in (0, \delta)$, (53) holds as a strict inequality. The claim follows from the hypothesis that (18) holds as a strict inequality for each i, combined with the assumption that f_i is continous and the fact $\lim_{s \downarrow 0} \phi_i(x_i, s) = \psi_i(x_i)$, due to Eq. (36) and the continuity of γ_i (Lemma 6.d). Thus, pick a sufficiently small $s \in (0, \delta)$ so that (i) $x_{-i} > 0 \Rightarrow s_i < x_{-i}$ for each $-i$ and (ii) $s < \bar{s}_i$ for each i, which can be done due of the fact $\bar{s}_i > 0$. With Property (i), Eq. (34) holds, which coupled with (53) implies that s admits a peaceful equilibrium. With Property (ii), $\phi_i(x_i, s) > 0$ for each i, hence the output $f_i\left(x_i, \phi_i(x_i, s)\right) > 0$. Thus, since $s > 0$, the center's tax revenue at the peaceful equilibrium is strictly positive. By contrast, if the center does not commit to a tax rate, Proposition 4 says that the center gets zero revenue; the center also gets zero revenue if it commits to a tax rate that results in wars. ■

Proposition 6 The center's revenue from a peaceful equilibrium, with a tax rate s, equals

$$R(s) := s\left(f_1(x_1, \phi_1(x_1, s)) + f_2(x_2, \phi_2(x_2, s))\right),$$

which is continuous in s, as f_i is continuous by assumption and ϕ_i continuous by Eq. (36) and Lemma 6. Thus, the maximum of R on the compact set $[0, 1-x_1-x_2]$ exists. By Proposition 5, the maximum value of R is strictly positive. Hence $s = 0$, yielding zero revenue, is not a maximum. Nor is $s = 1-x_1-x_2$ a maximum: Given this s, a region i's gross expected payoff become either $x_i\,(f_i(x_i,e_i)+y_{-i})$ when it pillages $-i$, or $(x_1+x_2)f_i(x_i,e_i)$ in case of peace, or $(1-x_{-i})\,(f_i(x_i,e_i)+y_{-i})$ if it is pillaged; consequently, as in the proof of Proposition 4, internal warfare is the only equilibrium outcome, rendering zero revenue. ■

G Proof of Remark 3

G.1 Calculating the Optimal Decentralized System

By Eq. (37), $D_2 f_i(x_i,e_i) = 4/(e_i+1)$ for both regions i. Hence the solutions ψ_i, η_i and η_i for the problems in (7), (8) and (35) are identical for both i, so we suppress their subscript i:

$$\begin{aligned}
\psi(x_i) &= 4/c-1,\\
\eta(x_i) &= (4x_i/c-1)^+,\\
\phi(x_i,s) &= (4(1-s)/c-1)^+, \qquad (54)
\end{aligned}$$

where the first line uses the assumption $c<4$ in the remark. Thus,

$$\begin{aligned}
f_i(x_i,\psi(x_i)) &= \ln(x_i+1)+4\ln(4/c),\\
f_i(x_i,\eta(x_i)) &= \ln(x_i+1)+4\left(\ln(4/c)+\ln x_i\right)^+, \qquad (55)\\
f_i(x_i,\phi(x_i,s)) &= \ln(x_i+1)+4\left(\ln(4/c)+\ln(1-s)\right)^+, \qquad (56)\\
c\psi(x_i) &= 4-c,\\
c\eta(x_i) &= (4x_i-c)^+, \qquad (57)\\
c\phi(x_i,s) &= (4(1-s)-c)^+. \qquad (58)
\end{aligned}$$

Since $f_i(x_i,\eta(x_i))$ is continuous in x_i, and it is differentiable except when x_i equals $c/4$, the social surplus generated by a fully decentralized power allocation $(x_1, 1-x_1)$ is continuous in x_1, and it is differentiable except when x_1 equals $c/4$ or $1-c/4$ (i.e., $x_2 = c/4$). Thus, in calculating the social surplus of full decentralization there are at most four possible cases:

i. $x_1 < c/4$ and $x_1 \le 1-c/4$ (i.e., $x_2 \ge c/4$);

ii. $x_1 < c/4$ and $x_1 > 1-c/4$ (i.e., $x_2 < c/4$);

iii. $x_1 \ge c/4$ and $x_1 \le 1-c/4$ (i.e., $x_2 \ge c/4$);

iv. $x_1 \geq c/4$ and $x_1 > 1 - c/4$ (i.e., $x_2 < c/4$).

Note that Cases (i) and (iv) are symmetric. Thus, given that the two regions are symmetric in their parameters, we need only to consider the social surpluses in the first three cases, respectively denoted by $S_{(\mathrm{i})}(x_1)$, $S_{(\mathrm{ii})}(x_1)$ and $S_{(\mathrm{iii})}(x_1)$, which are, by Eqs. (55) and (57),

$$\begin{aligned} S_{(\mathrm{i})}(x_1) &= \ln(x_1+1) + \ln(2-x_1) + 4\ln(4/c) + 4\ln(1-x_1) - 4(1-x_1) + c, \quad (59)\\ S_{(\mathrm{ii})}(x_1) &= \ln(x_1+1) + \ln(2-x_1), \quad (60)\\ S_{(\mathrm{iii})}(x_1) &= \ln(x_1+1) + 4\ln(4/c) + 4\ln x_1 - 4x_1 + c \\ &\quad + \ln(2-x_1) + 4\ln(4/c) + 4\ln(1-x_1) - 4(1-x_1) + c \\ &= 8\ln(4/c) + 2c + \ln(x_1+1) + 4\ln x_1 + \ln(2-x_1) + 4\ln(1-x_1) - 4. \quad (61) \end{aligned}$$

Claim 1 *If* $0.36 < c < 3.64$ *then the maximizer of* $S_{(\mathrm{i})}$ *on* $[0, \min\{c/4, 1-c/4\}]$ *is approximately equal to 0.089708, with maximum value*

$$S^*_{(\mathrm{i})} \approx 4\ln(4/c) + c - 3.2839617. \quad (62)$$

Proof Since $\frac{d}{dx_1}S_{(\mathrm{i})}(x_1) = 1/(x_1+1) - 1(2-x_1) - 4/(1-x_1) + 4$ is strictly decreasing in x_1, $S_{(\mathrm{i})}$ is strictly concave, hence if its domain contains the solution for $\frac{d}{dx_1}S_{(\mathrm{i})}(x_1) = 0$ then the solution is the unique maximum of $S_{(\mathrm{i})}$ on the domain. The equation $\frac{d}{dx_1}S_{(\mathrm{i})}(x_1) = 0$ is

$$4x_1^3 - 2x_1^2 - 11x_1 + 1 = 0,$$

whose solution exists and is approximately 0.089708, which belongs to the domain of $S_{(\mathrm{i})}$ by the hypothesis $0.36 < c < 3.64$ of this claim. Hence by Eq. (59)

$$\begin{aligned} S^*_{(\mathrm{i})} &\approx \ln(1.089708 \times 1.910292) + 4\ln(4/c) + 4\ln(0.910292) - 4 \times 0.910292 + c \\ &\approx 4\ln(4/c) + c + 0.7331659 + 4 \times (-0.0939899) - 3.641168, \end{aligned}$$

so the claim follows. ■

Claim 2 *If* $1/2$ *belongs to the interval* $[\min\{c/4, 1-c/4\}, c/4]$ *then* $S_{(\mathrm{ii})}$ *restricted on this interval attains to its maximum value* $S^*_{(\mathrm{ii})} = 2\ln(3/2)$ *at* $x_1 = 1/2$*; else* $S_{(\mathrm{ii})}$ *restricted on this interval is less than the maximal value of* $S_{(\mathrm{i})}$ *or* $S_{(\mathrm{iii})}$*.*

Proof Since $\frac{d}{dx_1}S_{(\mathrm{ii})}(x_1) = 1/(x_1+1) - 1(2-x_1)$ is strictly decreasing in x_1, $S_{(\mathrm{ii})}$ is strictly concave. Thus $S_{(\mathrm{ii})}$, without the restriction within the interval $[\min\{c/4, 1-c/4\}, c/4]$, attains to its unique maximum on $[0,1]$ at the solution for $\frac{d}{dx_1}S_{(\mathrm{ii})}(x_1) = 0$, i.e., $x_1 = 1/2$. Hence follows the claim. ■

Claim 3 *If* $1/2$ *belongs to the interval* $[c/4, \max\{c/4, 1-c/4\}]$ *then* $S_{\text{(iii)}}$ *restricted on this interval attains to its maximum at* $x_1 = 1/2$*, with maximum value*

$$S^*_{\text{(iii)}} = 8\ln(4/c) + 2c + 2\ln 3 - 10\ln 2 - 4; \tag{63}$$

else $S_{\text{(ii)}}$ *restricted on this interval is less than the maximal value of* $S_{\text{(i)}}$ *or* $S_{\text{(ii)}}$*.*

Proof Since $\frac{d}{dx_1}S_{\text{(iii)}}(x_1) = 1/(x_1+1) + 4/x_1 - 1(2-x_1) - 4/(1-x_1)$ is strictly decreasing in x_1, $S_{\text{(ii)}}$ is strictly concave, and the proof of this claim is analogous to that of Claim 2, and by Eq. (61) we have

$$S^*_{\text{(iii)}} = S_{\text{(iii)}}(1/2) = 8\ln(4/c) + 2c + \ln(3/2) + 4\ln(1/2) + \ln(3/2) + 4\ln(1/2) - 4.$$

Hence follows the claim. ■

Thus, when $0.36 < c < 3.64$, the social-surplus maximum among fully decentralized systems is either the symmetric allocation $(1/2, 1/2)$ or the asymmetric one that gives approximately 0.089708 power to one of the regions, and the rest to the other region. That proves part (a) of the remark.

G.2 A Peaceful System, totally Centralized

Let us consider the power allocation $(0, 0)$, giving all power to the center. We shall prove that this totally centralized allocation admits a peaceful equilibrium, with the center choosing a revenue-maximizing tax rate that belongs to the interval $(0.36, 0.362)$ when the effort cost $c = 1.45$. By continuity our demonstration is robust to small perturbations of c.

Claim 4 *The power allocation* $(0, 0)$*, coupled with any tax rate* $s \in [0, 1]$ *chosen by the center, admits a peaceful equilibrium.*

Proof Mimic the proof of Lemma 9 and prove that Ineq. (53) holds for each $i \in \{1, 2\}$. ■

The next claim proves Part (b) of the remark.

Claim 5 *Given power allocation* $(0, 0)$*, the center's expected revenue is higher when it commits to some tax rate than when it does not.*

Proof If the center does not commit then, by Proposition 4, after the outputs are produced, the center will extract all of them from the powerless regions, hence each region exerts zero effort, which coupled with zero power in the region renders zero output and hence zero tax revenue to the center. Whereas, the center can commit to a tiny positive tax rate so that each region exerts some positive amount of efforts, which result in a positive amount of outputs thereby positive amount of tax revenues for the center. ■

Claim 6 *If* $c = 1.45$ *and the power allocation is* $(0,0)$*, then the center's revenue-maximizing tax rate belongs to the interval* $(0.36, 0.362)$*.*

Proof Given the power allocation $(0,0)$, the center can choose any tax rate s from $[0,1]$. However, any tax rate above 0.6375 is strictly suboptimal to the center, because $s > 0.6375$ means $1 - s < 1.45/4$ and hence by Eq. (54) each region will exert zero effort, rendering zero revenue for the center; whereas a sufficiently small but positive tax rate would generate some revenue. Thus, there is no loss of generality to restrict attention to those tax rates $s \in [0, 0.6375]$. Any such s gives the center an expected revenue equal to, by Eq. (56),

$$\begin{aligned} R(s) &= s\left(\ln(x_1+1) + \ln(x_2+1) + 8\ln(4/c) + 8\ln(1-s)\right) \\ &= s\left(8\ln(4/c) + 8\ln(1-s)\right) \end{aligned}$$

since $x_1 = x_2 = 0$ given the power allocation. Hence

$$\frac{1}{8}\frac{d}{ds}R(s) = \ln(4/c) + \ln(1-s) - s/(1-s),$$

which is strictly decreasing in s. Thus, if a solution for $\frac{d}{ds}R(s) = 0$ exists in $[0, 0.6375]$ then it is the unique revenue maximizer. With $c = 1.45$,

$$\begin{aligned} \frac{1}{8}\left.\frac{d}{ds}R(s)\right|_{s=0.36} &= \ln(4/1.45) + \ln(0.64) - 0.36/0.64 \approx 0.00594, \\ \frac{1}{8}\left.\frac{d}{ds}R(s)\right|_{s=0.362} &= \ln(4/1.45) + \ln(0.638) - 0.362/0.638 \approx -0.0021. \end{aligned}$$

Hence the claim follows from the intermediate-value theorem. ■

G.3 Centralization Outperforms Decentralization

Now we shall demonstrate that the totally centralized system $(0,0)$, with the selfish center choosing a revenue-maximizing tax rate, generates a larger social surplus than even the best among the decentralized systems when the effort cost $c = 1.45$. The demonstration is again robust to small perturbations of c.[21]

[21] One can prove that the totally centralized allocation generates larger expected social surplus than the optimal decentralized system does for any $c \in [1.45, 1.55]$.

Claim 7 *If $c = 1.45$ then the maximum social surplus among fully decentralized systems is equal to $S^*_{\text{(iii)}}$.*

Proof With $c = 1.45$, $1/2$ belongs to the domain $(1.45/4, 1 - 1.45/4)$ for $S_{\text{(iii)}}$, hence by Claims 2 and 3, $S^*_{\text{(iii)}} \geq S^*_{\text{(ii)}}$. Thus, the maximum social surplus among fully decentralized systems is $\max\left\{S^*_{\text{(i)}}, S^*_{\text{(iii)}}\right\}$. By Eqs. (62) and (63) and $c = 1.45$,

$$\begin{aligned}
S^*_{\text{(i)}} - S^*_{\text{(iii)}} &\approx 4\ln(4/c) + c - 3.2839617 - (8\ln(4/c) + 2c + 2\ln 3 - 10\ln 2 - 4) \\
&= -4\ln(4/c) - c - 3.2839617 - 2\ln 3 + 10\ln 2 + 4 \\
&\approx -4(1.014731) - 1.45 - 3.2839617 - 2(1.098612) + 6.931472 + 4 \\
&\approx -0.058638.
\end{aligned}$$

Hence $S^*_{\text{(i)}} < S^*_{\text{(iii)}}$, as claimed. ■

Claim 8 *If $c = 1.45$ then the social surplus generated by the totally centralized power allocation $(0, 0)$ is larger than $S^*_{\text{(iii)}}$.*

Proof By Claim 6, the tax rate $s < 0.362 < 0.6375 = 1 - c/4$, hence the social surplus generated by the power allocation $(0, 0)$ is equal to, by Eqs. (56) and (58),

$$S_0 = 2\left(4\ln(4/c) + 4\ln(1 - s) - 4(1 - s) + c\right),$$

which by Eq. (63) implies

$$\begin{aligned}
S_0 - S^*_{\text{(iii)}} &= 2\left(4\ln(4/c) + 4\ln(1 - s) - 4(1 - s) + c\right) - (8\ln(4/c) + 2c + 2\ln 3 - 10\ln 2 - 4) \\
&= 8\ln(1 - s) - 8(1 - s) - 2\ln 3 + 10\ln 2 + 4 \\
&\approx 8.734247 + 8\ln(1 - s) - 8(1 - s).
\end{aligned}$$

Since the expression $8\ln(1 - s) - 8(1 - s)$ on the right-hand side is strictly decreasing in s, and $s < 0.362$ by Claim 6, the last displayed expression is greater than

$$\begin{aligned}
&8.734247 + 8\ln(1 - 0.362) - 8(1 - 0.362) \\
\approx\ &8.734247 + 8(-0.449417) - 8(0.638) \\
=\ &0.034911 \ > \ 0. \quad ■
\end{aligned}$$

By Claims 7 and 8, $S_0 > \max\left\{S^*_{\text{(i)}}, S^*_{\text{(ii)}}, S^*_{\text{(iii)}}\right\}$. That proves Part (c) of the remark.

References

[1] Daron Acemoglu, Philippe Aghion, Claire LeLarge, John van Reenen, and Fabrizio Zilibotti. Technology, information, and the decentralization of the firm. *Quarterly Journal of Economics*, 122(4):1759–1799, 2007. 1

[2] Daron Acemoglu, Georgy Egorov, and Konstantin Sonin. Coalition formation in non-democracies. *Review of Economic Studies*, 75:987–1009, 2008. 1

[3] Daron Acemoglu, Mikhail Golosov, Aleh Tsyvinski, and Pierre Yared. A dynamic theory of resource wars. *Quarterly Journal of Economics*, 127:283–331, 2012. 1

[4] Daron Acemoglu and James Robinson. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press: Cambridge, 2006. 18

[5] Ricardo Alonso, Wouter Dessein, and Niko Matouschek. When does coordination require centralization? *American Economic Review*, 98:145–179, 2008. 1

[6] J. Yannis Bakos. Reducing buyer search costs: Implications for electronic marketplace. *Management Science*, 43(12):1676–1692, 1997. 9

[7] Sandeep Baliga and Tomas Sjötröm. Decentralization and collusion. *Journal of Economic Theory*, 83:196–232, 1998. 1

[8] Sandeep Baliga and Tomas Sjötröm. The strategy of manipulating conflict. *American Economic Review*, 102(6):2897–2922, 2012. 1

[9] Sandeep Baliga and Tomas Sjötröm. Bargaining and war: A review of some formal models. *Korean Economic Review*, 29:235–266, 2013. 1

[10] Pranab Bardhan and Dilip Mookherjee. Capture and governance at local and national levels. *American Economic Review*, 90(2):135–139, 2000. 1

[11] Timothy Besley and Stephen Coate. Centralized versus decentralized provision of local public goods: A political economy approach. *Journal of Public Economics*, 87:2611–2637, 2003. 1

[12] Julian Birkinshaw. Strategies for managing internal competition. *California Management Review*, 44(1):21–38, 2001. 7, 11

[13] Olivier Blanchard and Andrei Shleifer. Federalism with and without political centralization: China versus Russia. *IMF Staff Papers*, 48:171–179, 2001. 1

[14] Hongbin Cai and Daniel Treisman. State corroding federalism. *Journal of Public Economics*, 88:819–843, 2004. 1

[15] Francesco Caselli, Massimo Morelli, and Dominic Rohner. The geography of interstate resource wars. *Quarterly Journal of Economics*, 130:267–315, 2015. 1

[16] Jared Diamond. *Guns, Germs, and Steel: The Fates of Human Societies*. W. W. Norton, 1997. 4

[17] Edward Gibbon. *The Decline and Fall of the Roman Empire*. Everyman's Library, 1910. First published between 1776 and 1788. 5

[18] Friedrich A. Hayek. The use of knowledge in society. *American Economic Review*, 35(4):519–530, September 1945. 13

[19] Erik Hedlund and Christina Frick. The fear of cannibalization: A quantitative study of Polarn O. Pyret's multi-channel strategy. Stockholm School of Economics Bachelor Thesis, Spring 2011. 9

[20] Herodotus. *The Histories*. Penguin Books: Baltimore, 1954. Translated by Aubrey de Sélincourt; with an introduction and notes by A. R. Burn. 1, 6

[21] Thomas Hobbes. On man, being the first part of Leviathan. In Charles W. Eliot, editor, *French and English Philosophers: The Harvard Classics*, volume 34, pages 311–413. P. F. Collier & Son: New York, 1910. 1, 4.1, 15, 5.2

[22] Matthew Jackson and Massimo Morelli. Political bias and war. *American Economic Review*, 97(4):1353–1373, 2007. 1

[23] Matthew Jackson and Massimo Morelli. The reasons for wars: An updated survey. In *Handbook on the Political Economy of War*, pages 34–57. Cheltenham/Northampton, Edward Elgar, 2011. 1

[24] James S. Jordan. Power and Legitimacy in Pillage Games. Mimeo, Pennsylvania State University, 2006. 1

[25] James S. Jordan. Pillage and property. *Journal of Economic Theory*, 131:26–44, 2006. 1

[26] James S. Jordan. Power and efficiency in production pillage games. *Review of Economic Design*, 13(3):171–193, 2009. 1

[27] Arturs Kalnins. An empirical analysis of territorial encroachment within franchised and company-owned branded chains. *Marketing Science*, 23(4):476–489, 2004. 8

[28] Peter Klibanoff and Jonathan Morduch. Decentralization, externalities, and efficiency. *Review of Economic Studies*, 62:223–247, 1995. 14

[29] Peter Klibanoff and Michael Poitevin. A Theory of (De)centralization. Mimeo, April 2013. 1

[30] Alan Kraus and Amir Rubin. Reducing managers' incentives to cannibalize: Managerial stock options when shareholders are diversified. *Journal of Financial Intermediation*, 19(4):439–460, 2010. 12

[31] John Locke. *Second Treatise of Government*. Harlan Davidson: Arlington Heights, IL, 1982. 1

[32] Eric Maskin, Yingyi Qian, and Chenggang Xu. Incentives, information, and organization form. *Review of Economic Studies*, 67:359–378, 2000. 1

[33] Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, March 2002. A

[34] Paul Milgrom and Chris Shannon. Monotone comparative statics. *Econometrica*, 62(1):157–180, 1994. C

[35] Joel Mokyr. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton University Press, 2002. 3

[36] Dilip Mookherjee. Decentralization, hierarchies and incentives: A mechanism design perspective. *Journal of Economic Literature*, 64:367–390, 2006. 13

[37] Massimo Morelli and Dominic Rohner. Resource concentration and civil wars. *Journal of Development Economics*, 117:32–47, 2015. 1

[38] Roger B. Myerson. Federalism and incentives for success of democracy. *Quarterly Journal of Political Science*, 1(1):3–23, 2006. 1

[39] Roger B. Myerson. Democratic decentralization and economic development. In Célestin Monga and Justin Yifu Lin, editors, *The Oxford Handbook of Africa and Economies: Volume 1: Context and Concepts*. Oxford University Press, 2015. 3

[40] Roger B. Myerson. Moral hazard in high office and the dynamics of aristocracy. *Econometrica*, 83(6):2083–2126, 2015. 7.2

[41] Herakles Polemarchakis and Colin Rowat. An Exchange Economy with Pillage. Draft, 2007. 1

[42] Raúl Ponce-Rodríguez, Charles Hankla, Jorge Martinez-Vazquez, and Eunice Heredia-Ortiz. Rethinking the political economy of decentralization: How elections and parties shape the provision of local public goods. International Center for Public Policy Working Paper 12-27, August 2012. 1

[43] Jean-Jacques Rousseau. *The Social Contract*. Penguin Books: New York, 1968. Translated by Maurice Cranston. 1

[44] Walter Scheidel. From the 'Great Convergence' to the 'First Great Divergence': Roman and Qin-Han state formation and its aftermath. In Walter Scheidel, editor, *Rome and China: Comparative Studies on Ancient World Empires*. New York: Oxford University Press, 2009. 2

[45] Joseph Stiglitz. *Whither Socialism?* The MIT Press: Cambridge, MA, 1994. 13

[46] Thucydides. *The Peloponnesian War*. Penguin Books: Baltimore, 1954. Translated by M. I. Finley. 6