Elite Persistence, Power Struggles and Coalition Dynamics*

Xinyu Fan†

January 21, 2018

Abstract

Why do social and political hierarchies often prove extremely difficult to eradicate? This paper studies the emergence and persistence of elites after power struggles. Players use their power to eliminate others and split resources. Players can also strategically give away power, i.e., burn power to invite new alliances or buy off key members to avoid expulsion. In equilibrium, the weak players cede power to the strong in order to deter regime changes. This ensures the survival of the weak, at the cost of increased inequality. We characterize the equilibrium power structures and study their robustness. Interestingly, perturbations on equilibrium structures follow the Iron Law of Oligarchy: power often ends up more concentrated to a few elite members, regardless of the immediate effects of power shifts. The model helps to explain why revolutions that aim to install social equality quickly reproduce the same type of hierarchies that the revolutionaries sought to destroy. We also discuss how elite persistence is influenced by economic development, external threats and international cooperation.

JEL Codes: D72, D74, O43, P41

Key Words: Power Struggles, Power Burning, Coalition Formation, Elite Persistence, Oligarchy

*The latest version of the paper is available at: www.xinyufanecon.com
The author is deeply indebted to Simon Board, as well as Michael Chwe, Moritz Meyer-ter-Vehn, Marek Pycia, Tomasz Sadzik and Daniel Treisman for their supervison, support and continued guidance. The author is also grateful to Omer Ali, Nick Doran, Hugo Hopenhayn, Jay Lu, Zhuoran Lu, John Nye, Ichiro Obara, Konstantin Sonin, Lingwei Wu, Chenggang Xu, Lixin Colin Xu, Feng Yang, Pu Zhao, and participants in the 2017 Ronald Coase Institute Workshop of Institutional Analysis, the 5th International Workshop on Economic Analysis of Institutions at Xiamen University, and the UCLA theory pro-seminar.

†Department of Economics, University of California, Los Angeles. Email: fanxy@ucla.edu.

"*Even when the discontent of the masses culminates in a successful attempt to deprive the bourgeoisie of power, this is ... effected only in appearance; always and necessarily there springs from the masses a new organized minority which raises itself to the rank of a governing class.*"

Robert Michels, *Political Parties: A sociological study of the oligarchical tendencies of modern democracy,* 1915, p.390.

1 Introduction

Why do social and political hierarchies often prove extremely difficult to eradicate? In many cases, revolutions that aim to install social equality quickly reproduce the same type of hierarchies that the revolutionaries sought to destroy. In George Orwell's terms, it turns out "some animals are more equal than others". The Bolshevik Revolution, rather than establishing the rule of the proletariat, quickly degenerated into an oligarchy led by Lenin, and later a dictatorship of Stalin for 30 years (Pipes 2011). Likewise, the French revolutionaries, after overthrowing their king, soon produced an emperor (De Tocqueville 1856). The durability of unequal power distribution has usually been attributed to culture - the "serf mentality" of Russians (Havel 1985) or Confucian values in Asia (Zhang et.al 2005) - or to the rapacious energy of dominant individuals (Hobbes 1651). The seminal paper of Acemoglu, Egorov, and Sonin (2008, hereafter AES) points out that rationality alone can explain the break-down of hierarchies: the alliance of several weak members can expel the strong in power struggles and sustain the rule, as long as there are no further power struggles within the alliance. However, in AES, players have fixed power levels, thus cannot respond to elimination threats by accepting a lower-ranked position, like Nikolai Bukharin of the Soviet Union after 1929 (Gregory 2013); or by pledging allegiance to others, like the Sénat Conservateur to Napoleon Bonaparte after the coup of 18 Brumaire (Doyle 1989). Furthermore, it remains a puzzle why "always and necessarily there springs from the masses a new organized minority which raises itself to the rank of a governing class" (Michels 1915) after the old hierarchy falls.

The contribution of this paper is to explain the emergence of an elite group after power struggles and its persistence, by introducing power burning as a commitment device during power struggles. Power struggle is modeled as an iterative coalition formation process where a group of players uses their power to form alliances, eliminate others, and split fixed resources according to relative power if they survive the power struggles. In particular, players can strategically give away power, i.e. burn power, and then participate in power struggles to eliminate others and split resources. By doing so they can credibly invite new alliances or buy off key members in the coalition in order to avoid expulsion. In equilibrium, the weak players cede power to a few strong players who are able to enforce a smaller ruling circle, in order to deter regime changes. This ensures the survival of the weak, at the cost of increased inequality. We characterize the equilibrium structures in general and further study the robustness of such structures towards perturbations. Interestingly, perturbations on equilibrium structures follow the Iron Law of Oligarchy: power often ends up more concentrated to a few elite members, regardless of the immediate effects of power shifts. Thus the elite persist during ups and downs.

The emergence of elites results as follows. When a group of players form an alliance to eliminate others, they have to make sure the current alliance is safe from power struggles in the future. In particular, a strong player sometimes has to give away power to credibly assure his allies that he is not a future threat, in the effort of forming a smaller circle to split the rents. This, on the other hand, gives the excluded players an opportunity to buy off the strong players by ceding power preemptively: the strong players receive enough benefits from the power ceding thus are reluctant for further regime changes. That is, the flexibility of power adjustment serves as a commitment device for strong players to threaten to form a smaller ruling circle. In response, the players outside the ruling circle (outsiders) have to give away their power until some of the players in the smaller ruling circle (elites) prefers the status quo. As a result, power becomes more concentrated, and a hierarchy is established in equilibrium between the outsiders and the elites. After the elites establish their prestigious positions, any perturbations that make them stronger naturally centralize the power structure, while any perturbations that weaken their positions re-establish the incentives to eliminate the outsiders. Consequently, the outsiders cede power again to restore the prestige of the elites, and the hierarchy persists.

The model consists of a coalition of agents, who split a fixed rent of 1 according to their respective, endowed levels of power. For example, in a 4-agent coalition with power structure $(2, 4, 5, 10)$, the first agent (with power level of 2) claims rents of $\frac{2}{2+4+5+10} = \frac{2}{21}$, while the rest of the agents claim $\frac{4}{21}$, $\frac{5}{21}$ and $\frac{10}{21}$, respectively. If agents are unsatisfied with current rent distribution, they can engage in power struggles in order to attempt to change the structure. Power struggles are modeled as an iterative process of two meetings. The first meeting is called a *power burning* meeting, where each agent sequentially proposes a power adjustment scheme to a sub-coalition of agents (the alliance) to prepare for the coming struggles. We assume agents only adjust power downwards (eg. disarm an army, resign from office, migrate to a foreign country, etc.), and refer to this process as *power burning*. A burning scheme is enacted only with unanimous consent from the included agents. Otherwise no power is burnt and it is the next agent's turn to propose. The burning meeting continues until no agent nor sub-coalition wishes to further burn power.

The second meeting is an *elimination meeting*. Each agent sequentially proposes a new, and potentially smaller ruling coalition. The proposal passes if the coalition commands strict majority power and only with the unanimous consent of all agents within the coalition. Otherwise the proposal is rejected and it turns to the next agent to propose. If all proposals are rejected, the game ends and the rents are split between agents based on the current power structure. If a new coalition is formed, all excluded agents are eliminated (eg. retired, exiled, or executed). The power struggle cycles continue in the new coalition: first the burning meetings and then the elimination meeting. An equilibrium is reached when no further burning or eliminations take place. Agents collect payoffs in equilibrium according to their status. A surviving agent splits the rents proportional to his power; an eliminated agent receives a large and negative payoff.

We show that a sub-game perfect Nash Equilibrium (SPE) always exists for the power struggle game, and the equilibrium structure is unique given initial power structure and the sequence of moves (Theorem 1). In Section 3 and 4, we study the formation of equilibrium structure to show how hierarchies resurge in equilibrium and how the elites maintain their prestige during power perturbations.

We illustrate the main economic forces using the following example. In the power struggle game, agents seek to form smaller alliances for higher rents while watching out for eliminations. Recall the example of $(2, 4, 5, 10)$, which resembles a hierarchy with an elite agent holding power of 10, two medium agents holding power of 4 and 5, and a weak agent holding power of 2. Based on the rules of power struggles, the alliance of the non-elites, $(2, 4, 5)$ is strong enough to overthrow the elite $(2 + 4 + 5>10)$ in revolution and split the rent within the alliance.

Power struggles do not stop after the old hierarchy falls. Suppose the old elite is eliminated, in the new regime of $(2, 4, 5)$, a new alliance between agent with power 4 and 5 is powerful enough to eliminate the weak agent. However, the coalition $(4, 5)$ is not stable after the elimination of the weak, because the stronger agent with power 5 can now expel the agent with power 4 and become a dictator. Anticipating further power struggles, $(4, 5)$ is not a profitable deviation for the agent with power 4. The current structure $(2, 4, 5)$ is thus stable. This is the idea of self-enforcing coalitions in AES.

However, when agents are allowed to burn power, agents behave differently. First, during the fall of the old hierarchy, the agent with power 10, facing elimination, can burn up his power for survival to avoid the negative payoff brought by elimination, $(2, 4, 5, 10) \rightarrow (2, 4, 5, 0)$. Next, in the new regime, $(2, 4, 5, 0)$, the agent with power 5 could credibly commit to the alliance by burning power down to 4, thus inducing a post-burning alliance of $(4, 4)$. Such an alliance is not only rent-improving for both participants, but also safe from future power struggles because no individual is powerful enough to expel the other one. We call such alliances *stable improving* coalitions.

Anticipating the formation of the stable improving alliance, the excluded agent has to cede power for survival. He does so to the degree that at least one of the insiders gets the same rents in the current regime as in the alternative. In the example, the agent with power 2 burns power to 1, $(2, 4, 5, 0) \to (1, 4, 5, 0)$, so that strong agent (with power 5) is rent-indifferent between the current regime $\left(\frac{5}{1+4+5}\right)$ and the stable improving alternative $\left(\frac{4}{4+4}\right)$. The new power structure $(1, 4, 5, 0)$ is the equilibrium structure.

We characterize the equilibrium structure in Theorem 2. A coalition reaches equilibrium in power struggles either when there does not exists a strictly stable improving alternative, or when some early-moving powerful agents are able to deter the regime change. In the example, $(4, 4)$ is the unique weakly stable improving alternative. Therefore the structure $(1, 4, 5, 0)$ is in equilibrium. Power is more concentrated in the equilibrium structure than the initial structure $(2, 4, 5, 0)$. That is, by making the weak even weaker to accommodate the strong, a new hierarchy emerges to protect the weak from elimination.

In reality, the power structure after a hierarchy established usually stays centralized during perturbations, if not more so[1]. To see this in the model, recall the equilibrium structure in the previous example $(1, 4, 5)$ (here we ignore the agent with zero power as he does not participate in the rent distributions). Suppose a negative perturbation applies to the elite, $(1, 4, 5) \to (1, 4, 5 - \epsilon)$. Now the elite has an incentive to ally with the agent with power 4 and eliminate the weak because their alliance $(4, 4)$ is now strictly stable improving $(\frac{1}{2} > \frac{5-\epsilon}{1+4+5-\epsilon})$. To block this alliance, the weak again cedes power, which leads to the new equilibrium structure $(1 - \epsilon, 4, 5 - \epsilon)$: the elite gets to maintain his rent after the negative shocks. In Theorem 3, we show that the persistence of elite prestige holds in general. By the same logic, suppose a positive perturbation hits the weak, $(1, 4, 5) \to (1 + \epsilon, 4, 5)$. The alliance $(4, 4)$ again becomes strictly stable improving, which forces the weak to cede back the increased power. The stable structure recovers to $(1, 4, 5)$: the weak remains weak . In Theorem 4 we show that the trap of the weak holds in general.

On the other hand, suppose a positive perturbation hits the elite, $(1, 4, 5 + \epsilon)$, or a negative perturbation hits the weak $(1 - \epsilon, 4, 5)$, then the elite has unchecked power to eliminate the rest of the coalition $(5 + \epsilon > 1 + 4\ )$ and becomes a dictator. To avoid elimination, the

[1] Magaloni and Kricheli (2010) record the increasingly rampant one-party rules in modern times. Svolik (2012) and Kendall-Taylor and Frantz (2016) highlight the persistence of successors' power after the deaths of old rulers.

two other agents cede power completely by burning down to 0, and the equilibrium structure becomes $(0,0,5+\epsilon)$ and $(0,0,5)$, respectively. Therefore, a perturbation, positive or negative, has a bias towards centralization, and a small perturbation may collapse the power structure to dictatorship. Combining the persistence of the elite and the trap of the weak, perturbing an equilibrium structure exhibits the Matthew effect[2]: regardless of the initial direction of perturbation, the new equilibrium always has a weakly more centralized power structure. In the literature of organizational structures, it is also referred as the Iron Law of Oligarchy (Michels 1915).

Even though the model is abstract, the insights apply to many real life scenarios. For example, the founding of the Song Dynasty in imperial China captures all of the previous forces. In 960 AD, at the Bridge of Chen, a military general named Zhao Kuangyin of Zhou Dynasty was "recommended" by his lieutenants to take the throne and become the new emperor. His lieutenants put a yellow dragon robe, which represented imperial power, on General Zhao. The incident turned out to be a peaceful coup as the current emperor, 9-year-old Chai Zongxun, agreed to give away power. He stepped down and lived a peaceful life thereafter. Known as the Coup of the Dragon Robe, this encapsulates the change $(2,4,5,10) \to (2,4,5,0)$ (Paludan 1998). However, the power struggles did not stop when General Zhao became Emperor Zhao. Two years later, Emperor Zhao "encouraged" some of his old lieutenants to retire in a grand banquet. The lieutenants agreed with "the tears of joy and regret (of not volunteering to do so earlier)", ceding their military command back to the regime. The regime became more centralized, which captures the changes $(2,4,5) \to (1,4,5)$ (Tao Li 2004, Vol.1). After taking back the military control, Emperor Zhao further took advantage of a small mistake in local jurisdiction to centralize both administrative and jurisdiction control, and soon seized all the power to himself. This last episode captures $(1-\epsilon,4,5) \to (0,0,5)$ (Tao Li 2004, Vol.14).

In Section 5, we apply our framework to consider two real-life scenarios and provide several testable predictions for power struggles and their consequences. The first scenario focuses on the tradeoff between economic development and extraction. While an intuitive tradeoff exists between efficiency and equality, power struggles introduce additional concerns of survival for the rulers. We show in Proposition 2 that, when the initiation of a power struggle is a policy choice for the leadership, a ruler with weak control over the society may support *more* economic development, because he might not succeed in the extractions

[2] The Matthew Effect is first proposed by Merton (1968) to describe the accumulated advantages in scientific research. The main idea is best summarized by the phenomenon that the rich becomes richer, and the poor becomes poorer.

through initiating a power struggle. That is, a weak ruler can be forced to focus on economic decentralization instead of extraction for fear of losing power in the extractions, which leads to a weaker ruler and the cycle continues. Correspondingly, we predict a far-sighted ruler should adopt a "sow and reap" strategy when he is strong: aim for economic development for some time, then initiate power struggles before the potential opponents gets too powerful, and resume economic development afterwards.

The second scenario is the domestic conflict resolution in presence of international influence. Intuitively, imminent foreign threats (Nazis on the Soviet Union, Mongolian riders at the gate, etc.) soften domestic conflicts because a united domestic power stands more chances against the threat. However, when external forces are no longer a threat, domestic coalitions resume their power struggles. We summarize the result in Proposition 3 and suggest the power struggles in post-war military leaderships could be available for empirical tests. On the other hand, Proposition 4 shows that an agent to be expelled can make himself indispensable in the power balance, by inviting a strong foreign force to settle the disputes. As an illustration, we point to the regime changes in Caribbean colonies where the increasingly powerful citizenry induced the elite to cede power back to the British Crown (Ashdown 1979, Carvalho and Dippel 2016).

1.1 Literature

Our discussion of power struggles is related to a number of literature. We echo the models of power struggles (Acemoglu Egorov Sonin 2006, 2008, 2012, Jandoc and Juarez 2013, Guimaraes and Sheedy 2017), in that participants care not only about the immediate impact of power structure changes, but also the aftermath, i.e. the struggles after struggles[3]. Building on AES, this is the first paper that allows agents to adjust their power to impose elimination threats or to respond to such threats. Furthermore, we adopt the novel notion of power burning to characterize equilibrium structures. Surprisingly, the endogenous power adjustment, which is expected to support richer equilibrium structures, turns out generating sharper predictions of equilibrium dynamics than the exogenous power levels: The flexibility in power adjustment gives the elite more opportunities to strengthen their advantageous positions.

Our model is also among the first papers to combine the roles of power in rent distribution (Baron and Ferejohn 1989) and the roles of power in endogenously determining coalition

[3]The far-sight in elimination for resources are also seen in the game of truels, or three-person duels (Kilgour 1977, Amengual and Toral 2006). Three players form a circle, each of them points his gun to the "downstream" player, and has his head aimed by the "upstream" one. A balance of terror can be maintained where no player pulls the trigger, which shares the insight of far-sighted power struggles.

composition (Pycia 2012, Ray and Vohra 2015). The seminal work by Baron and Ferejohn (1989) argues the resource distribution in legislature depends on exogenous participation constraints of involving parties. And one has to accommodate the necessary alliances to rule. Bueno de Mesquita et al. (2003) extend the insight to political survival and argue that the ruler has to accommodate the interests of his smallest winning coalition. However, it remains unclear how the participation constraints and the sizes of winning coalitions are determined. That is, the resource allocation role of power is based on fixed coalition composition. Our paper provides a micro-foundation of how participation constraints and winning coalitions are endogenously formed in power struggles, thus linking the distribution role of power to the coalition formation role.

In the literature of elite persistence in organizations, especially non-democracies, Acemoglu and Robinson (2006, 2008) and Robinson (2012) emphasize the office-holding elites implement policies and investments that favor themselves in the future. The insights are also confirmed in empirical studies of democracies, in the context of United States Congress (Dal Bó, Dal Bó, and Snyder 2007). Their models focus on the competition between an existing elite and, essentially, a representative citizen. This paper adds to the discussion by showing how the elites emerge and persist from a generic coalitional structure. To the best of the author's knowledge, this is also the first paper to show that the persistence of elites can be established and sustained through power struggles and far-sighted coalition in a zero-sum game, without opportunities for elites to accumulate power through investments (eg. Robinson 2008).

Our model aims to provide a framework of regime changes without explicit assumptions on institutional details. In the classical literature about regime changes, a menu of established regime choices are usually prerequisites: Acemgolu and Robinson (2005, 2013) on democracy vs. dictatorship; Fearon (2011) on ballot vs. bullet, etc. Discussions based on established regimes provide clear-cut comparisons, but implicitly assume away the origin of institutions, i.e. where does the menu come from. On the other hand, there are growing attempts to discuss regime changes abstract from institutional details (eg. Aghion and Bolton 2003, Aghion, Alesina and Trebbi 2004, Gehlbach, Sonin and Svolik 2016). Following the latter approach, this paper contains no presumed forms of institutions, thus adding to the literature by showing how familiar institutional designs may emerge as equilibrium outcomes.

2 Model

This section details the set-up of the model and discusses the existence and uniqueness of the equilibrium after power struggles. Our model of power struggles provides opportunities for players to adjust their power and eliminate other players. In brief, players first participate in *power burning meetings* where they sequentially propose downward power adjustment schemes until no coalition wishes to burn power further. Players then participate in *elimination meetings* where they sequentially form alliances to eliminate unallied members in order to gain higher relative power. Equilibrium is reached when no players wish to further adjust power or to eliminate others.

2.1 Set-up

Players, power, coalitions. We denote the set of finite *agents* by N_o. Agents are individuals or representatives for groups. A *coalition*, $N = \{1, 2, ..., n\}$, is a group of agents, $N \subset N_o$. Each agent in the coalition is endowed with some power $0 \leq \gamma_i < \infty$. γ_i is discrete and exists on small grids $\{0, \epsilon, 2\epsilon, ...\}$[4]. Agents are labeled from weak to strong in the *power vector*, $\gamma_N = (\gamma_1, \gamma_2, ..., \gamma_n)$, $\gamma_1 \leq \gamma_2 \leq ... \leq \gamma_n$. The power of a coalition, N, is simply the sum of its members' powers, $\sum_{i \in N} \gamma_i$. Within coalition N, a sub-coalition $C \subset N$, where $\gamma_{i \in C} = \gamma_{i \in N}$ for all $i \in C$, is *winning* if $\sum_{i \in C} \gamma_i > \sum_{j \in N \setminus C} \gamma_j$. That is, a sub-coalition is winning if it commands strict majority power.

Players move sequentially in the game. An *agenda* $s \in S$ specifies the sequence of moves and is fixed in a game [5]. We refer to a particular agenda as *top-down* when the (initially) strongest agent always moves first, followed by the second strongest, all the way down the hierarchy to the least powerful agent. An agenda is imposed on individuals, but not on power structures.

Strategies. Agents take actions in two kinds of meetings, iteratively. The game starts with stage-j coalition N_j, where in the first stage $N_1 = N$, and proceeds in the following steps:

1. [**Burning meeting**]: The burning meeting takes place in rounds. The history is visible to all agents. In Round 1, each agent, i, sequentially makes a burning proposal, $0 \leq \rho_i^{j,1}\left(\gamma_{B_{i,1}}^{j,0} | h_{1,s_i}\right) \leq \gamma_{B_{i,1}}^{j,0}$, to agents in sub-coalition $B_{i,1} \subset N$. h_{1,s_i} denotes the history up to Round 1 according to agenda s before i burns. Then agents in $B_{i,1}$ cast their votes sequentially according to s. Each agent votes yes or no. The proposal is passed if and only if all agents included in the proposed coalition vote yes. If passed, the proposal is implemented and power is burnt accordingly; otherwise no power adjustments take place and we proceed

[4] We assume power is discrete to cut the game into a finite horizon for the existence of equilibrium. In real life, power is often discrete, such as the size of an army or the population of one's followers.

[5] We assume a fixed agenda because turbulent power struggles often take place in a short period of time, where the sequence of moves of are hard to change due to geographic limits, etc. More discussions follow in Section 6.

to the next agent. A round finishes after all agents have made their proposals. In Round 2, each agent continues to make burning proposals $0 \leq \rho_i^{j,2}\left(\gamma_{B_{i,2}}^{j,1}|h_{2,s_i}\right) \leq \gamma_{B_{i,2}}^{j,1}$, following s. The process continues until Round $t+1$ where no burning takes place: for all $i \in N_j$, $\rho_i^{j,t+1}\left(\gamma_{B_{i,t}}^{j,t}|h_{t,s_i}\right) = \gamma_{B_{i,t}}^{j,t}$, $t \geq 1$. That is, a burning meeting lasts for at least two rounds. When the burning meeting ends, the game proceeds to Step 2.

2. [**Elimination meeting**]: The elimination meeting also takes place in rounds. Following the same agenda s, skipping any eliminated agents or agents with zero power, an agent, i, makes a proposal of a winning sub-coalition $P_i^j \subset N_j$ as the new ruling coalition. Agents then cast their votes sequentially according to s. Each agent in P_i^j votes yes or no. The proposal is passed if and only if all agents in P_i^j vote yes. The game proceeds to Step 3 below if the proposal is accepted; otherwise it turns to the next agent s_{i+1} according to s to propose. If no individual proposal is accepted in a round, the status quo proposal $P^j = N_j$ is accepted automatically. As a result, an elimination meeting always ends in one round.

3. Agents not in the proposal are eliminated. If the size of winning coalition shrinks, i.e. $|P_i^j| < |N_j|$, we go to stage $j+1$, where $N_{j+1} = P_i^j$, and start from Step 1 again. If $|P_i^j| = |N_j|$, we go to Step 4.

4. Payoffs are assigned.

Payoffs. Staying in the coalition generates office-holding rents. The total rents to share are fixed and are normalized to 1. Agents receive rents proportional to their relative power if they survive in power struggles. For an agent who survives in coalition N with power structure γ, he receives payoffs $U_i(N, \gamma) = \frac{\gamma_i}{\sum_{i \in N} \gamma_i}$. An eliminated agent receives $U_i = -n$, which captures the large costs of losing offices.

During the elimination meetings, if multiple non-proposing agents have the same power levels, we assume they have equal probabilities to be included in an elimination proposal. Given the large and negative payoff of being eliminated, an agent does not gamble for probabilistic inclusion in the stable coalition[6].

Assumption 1. Non-proposing agents with same power levels have equal probabilities to be included in an elimination proposal.

We also apply the following tie-breaking assumptions when an agent receives the same rents or has the same power level with another agent in elimination meetings:

Assumption 2. Agents always prefer a coalition where they burn less power when receiving

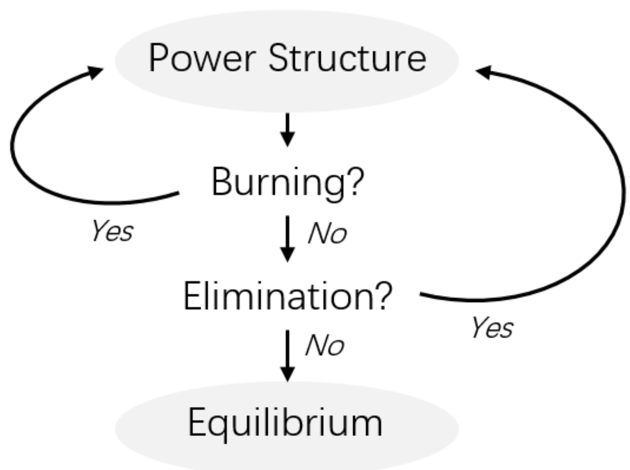
[6]For example, suppose there is a coalition with power structure $(3, 3, 3)$ and the agenda is top-down. Any two agents can form a winning coalition and eliminate the remaining agent. When agent 1 chooses his alliance, he is indifferent with choosing agent 2 and 3, thus assigning probability of $\frac{1}{2}$ to include either one. Both agents 2 and 3 face negative expected utility $\frac{1}{2}\left(\frac{1}{2} - n\right) < 0$. Under top-down agenda, the late-mover of the two, i.e. agent 3, has a better response by burning down to 0 and guarantees payoff of 0.

the same payoffs

Assumption 3. Agents always prefer a coalition formed with fewer meetings when the payoffs and power burnt are the same.

Agents stop attending meetings after they burn down to zero power. Assumption 2 captures the costly nature of burning power, for instance, the dismissal costs of a disarmament. Assumption 3 captures the meeting costs. These costs are usually small in scale relative to the rents, so we only consider them when breaking the ties. The lexicographic tie-breaking assumptions help to identify a unique stable structure in equilibrium.

Figure 1. Timing of the game



2.2 Existence and uniqueness of equilibrium

The solution concept is sub-perfect equilibrium (SPE). A coalition N^* with its power structure γ^* is in equilibrium if it survives the iterations of power burning and eliminations: $(N^*, \gamma^*) \in \phi(N, \gamma|s)$, where ϕ is the correspondence generated by the power struggle game, N is the initial coalition, γ is the initial power structure, and s is the agenda. Power struggles do not last forever because the power levels are discrete and thus the room for adjustment is limited. Therefore, power structures stabilize in finite numbers of meetings. Furthermore, with the tie-breaking assumptions, we are able to pick the unique equilibrium that aims to minimize power burning and the number of meetings.

Theorem 1. *There exists a SPE of the power struggle game. For a given agenda, s, and under Assumptions 1-3, the game induces a unique equilibrium power structure for initial coalition N with power structure γ, that is, $(N^*, \gamma^*) = \phi(N, \gamma|s)$.*

Proof. See Appendix. □

Theorem 1 shows that a unique outcome is reached after power struggles if the initial power structure and the agenda are fixed. Therefore, the direction of coalition dynamics is clear. Denote the initial power structure of a coalition as the *status quo*. By Theorem 1, if multiple equilibrium structures are reached starting from the same status quo, it is either due to the agenda differences, or the violations of at least one of the three assumptions. On the other hand, multiple status quos may converge to the same equilibrium structure. In the next section we discuss the impact of different agendas on equilibrium structures and the convergence of status quos, in detail.

Given the significant cost of being eliminated, the losing agents in the power struggles always cede all power to the rest of the coalition when they are about to be expelled, and forego the rights of claiming rents[7]. Therefore, we expect to see no elimination taking place on equilibrium path.

Corollary 1. *No elimination takes place on equilibrium path. In any equilibrium* $(N^*, \gamma^*) = \phi(N, \gamma|s)$*,* $N^* = N$*.*

Proof. See Appendix. □

This is in contrast with AES where agents not in the winning coalition are eliminated because they are unable to adjust their power. In our model, agents are allowed to give up their power to create new balances. Furthermore, the surviving agents may also give up power to initiate favorable regime changes, thus enriching the potential stable structures generated under power struggles. Surprisingly, such richness sharpens the predictions we can make towards equilibrium structures: we predict that the dynamics of equilibrium structures in response to external shocks often have a tendency to centralize. This will be the focus of our discussion in the next two sections.

Hereafter we suppress notations when there is no confusion about the coalition size or agenda referred to, and denote the equilibrium structure as $\gamma^* = \phi(\gamma)$. Also, we refer to "power struggles taking place" as the case where some power is burnt during the formation of equilibrium, that is, $\gamma \neq \phi(\gamma)$.

3 Three-Agent Coalitions

As motivations for the general characterization we present in Section 4, we analyze the equilibrium structure of a 3-agent coalition. The section shows how a hierarchical structure

[7]A good way of burning power for a politics-involving oligarch is perhaps selling your assets and emigrate to a foreign country, as Roman Abramovich, the current owner of Chelsea Football Club in Britain, did to his Russian conglomerate.

emerges in equilibrium, regardless of the initial power structure, and how the prestige of the elite persists through perturbations. A 3-agent coalition is also the minimal organization size that allows both competition and alliances.

3.1 Characterization

Consider a generic 3-agent coalition wherein agents have different power levels. We identify the agents as the weak (w), the medium (m) and the elite (e) by their power levels. They form a coalition with initial power structure $\gamma = (\gamma_w = a, \gamma_m = b, \gamma_e = c)$, where $0 < a < b < c$. When there are no power struggles, the *elite* receives rent of $\frac{c}{a+b+c}$; the *medium* receives $\frac{b}{a+b+c}$; and the *weak* receives $\frac{a}{a+b+c}$.

In power struggles, less powerful agents tend to form alliances to defeat the strong, while watching out for future power struggles within the alliance. When $a + b < c$, the alliance of the weak and the medium fails to compete with the elite. Facing the threat of elimination, they have to cede power to the elite completely, by burning down to 0, $(a, b, c) \to (0, 0, c)$. As a result, the equilibrium structure becomes $(\gamma_w = 0, \gamma_m = 0, \gamma_e = c)$. An extreme hierarchy is formed, which we will naturally refer to as a dictatorship.

When $a + b > c$, no agent is powerful enough to dictate, but any two-agent alliance is powerful enough to eliminate the third. A myopic weak agent may think that if he allies with the medium agent and they collectively expel the elite, then they both will receive higher payoffs in the new power structure $(\gamma_w = a, \gamma_m = b)$ because $\frac{a}{a+b} > \frac{a}{a+b+c}$ and $\frac{b}{a+b} > \frac{b}{a+b+c}$. However, a far-sighted weak agent understands that, after ousting the elite agent, the medium agent becomes dominant, and will proceed to eliminate the weak. The weak agent expects zero payoff from allying with the medium agent as a result. Anticipating that, the alliance of (a, b) is not rent-improving for the weak agent. Therefore, we see that no two-agent alliance is stable at the current power levels, which makes the original structure (a, b, c) stable in power struggles. This is the same as the idea of self-enforcing coalitions in AES.

Taking this idea one step further, suppose agents strategically burn power to prepare for the coming struggles. Each agent understands a two-agent alliance is stable only if they are of equal power. each agent receives rent of $\frac{1}{2}$, which is rent-improving for the two agents in the alliance. Meanwhile, the post-burning structure has to be powerful enough for the alliance to oust their opponents. For example, when $a < \frac{c}{2}$, an alliance $\left(\gamma_w', \gamma_m'\right) = (a, a)$, where the medium agent burns down to match power with the weak agent, fails to eliminate the elite. A rent-improving alliance between the medium and the elite agents is stable when the elite burns power to match with the medium, $\left(\gamma_m', \gamma_e'\right) = (b, b)$. There are other options of stable and rent-improving alliances, such as $\left(\gamma_m', \gamma_e'\right) = (b - \epsilon, b - \epsilon)$ for small ϵ. However, the tie-breaking assumption in Section 2.1 selects the outcome with the least power burnt, i.e. (b, b).

Anticipating the formation of (b, b) and his consequent elimination, the weak agent adjusts power to block the alliances. The weak agent does so by burning power until the elite enjoys same rents in the current coalition as in the alternative ($\frac{1}{2}$). As a result, the weak agent burns to $c - b$[8], which turns the power structure to $\gamma' = \left(\gamma'_w = c - b, \gamma'_m = b, \gamma'_e = c\right)$. The structure is in equilibrium because no agent wishes to further burn power or eliminate other agents. As a result, the flexibility in power adjustment provides credible threats for the elite to burn power, which leads to the weak agent burning power.

For any agenda, we can follow the practice above to look for alliances that are rent-improving, powerful enough to dominate in current power struggles, and are stable in future ones. If such an alliance exists, agents not in the alliance have to cede power for survival. For a 3-agent coalition, we completely characterize the equilibrium structure as follows:

Proposition 1. *Starting from a generic power structure $\gamma = (\gamma_w = a, \gamma_m = b, \gamma_e = c)$, $0 < a < b < c$, the equilibrium structure converges to one of the three forms:*

1. *a dictatorship: $\gamma^* = (0, 0, c)$ when $a + b < c$;*
2. *a duopoly: $\gamma^* = (a, a, 0)$ when $a + b \geq c$, $a > \frac{c}{2}$, $s = (m, w, e)$ or $s = (w, m, e)$;*
3. *an oligarchy: $\gamma^* = (c - b, b, c)$ otherwise.*

Proof. See Appendix. □

As the proposition shows, when the initial structure features an unbalanced power, the society necessarily falls into a dictatorship, without external forces to enforce rules or commitments. The other social classes fail to collectively compete with the elite, and must rally behind one agent for survival. This resembles the *established autocrats* in Svolik (2012), when the autocrats have no competitors. The complete power ceding from the rest of the coalition to the elite is also consistent with Acemoglu and Robinson's (2017) recent study, which shows regimes turn despotic when the power of the state is significantly stronger than the power of society.

On the other hand, when the initial power distribution is relatively equal, the equilibrium structure is an oligarchy. In an oligarchy, the rents received by the weak is irrelevant of the initial power he possesses. Instead, it is ironed down by the power difference between the medium and the elite agents. Power ironing also suggests that a variety of initial structures ends up with the same equilibrium, which may explain why only a few common forms of organizational structures are observed across geographic locations, cultures, and initial

[8] Recall that we are in the case where $a + b > c$, thus $c - b \geq 0$.

power endowments. This is in contrast with a classic paradigm of regime changes where the small collectively overthrow the strong and rule without concern of future power changes. During the formation of an oligarchy, the alliance between the less powerful (the weak and the medium) fails to compete with the alliance between the powerful (the medium and the elite). This is not because they are not powerful enough to overthrow the elite, but it is because of the stability concerns in the future. In this sense, power burning secures wider participation at the cost of power centralization.

The balance between the elite and the alliance of the rest of the coalition in oligarchies resembles the *contested autocracies* in Svolik (2012) where the autocrat's power is delicately balanced with other players. A natural question is whether and how the balance of power may break down and degenerate to unchecked power, i.e. the transition from contested to established autocracies. We address this question in Section 3.2 below.

Remarks on agenda. When agents have fixed power levels (as in AES), the formation of equilibrium structures is independent of agenda. However, as Proposition 1 suggests, when burning power is possible, the formation of equilibrium does depend on agenda[9], because the sequence of moves determines which alternate alliance can be formed first. As Olson (2000, p.xii) writes, the opportunities to consolidate authority "...occurs at times that Polish reformer Leszek Balcerowicz refers to as *moments of extraordinary politics*." Furthermore, the same initiate power structure could lead to drastically different equilibrium structures. Starting from the same initial power structure (a, b, c) when $a > \frac{c}{2}$, two alliances are stable and rent-improving: $\left(\gamma_w^{\prime}, \gamma_m^{\prime}\right) = (a, a)$ and $\left(\gamma_m^{\prime}, \gamma_e^{\prime}\right) = (b, b)$. When the elite moves last, i.e. $s = (m, w, e)$ or (w, m, e), the weak and the medium agent preempt to form the alliance of (a, a), which forces the elite agent to completely cede power. The equilibrium structure becomes $(a, a, 0)$. However, in all other agendas, for example, $s = (w, e, m)$, after the medium agent burns power to a, the elite can react by mimicking the first-mover and also burn down to a. This leads to a new stable structure, (a, a, a). The tie-breaking assumption suggests that the last-mover should yield, thus blocking the alliance of $\left(\gamma_w^{\prime}, \gamma_m^{\prime}\right) = (a, a)$ in the first place. This illustrates the survival concerns of the elite. When the initial power structure is highly skewed, the elite is always free of survival concerns. However, when the initial power structure is relatively equal, the elite could continue to expropriate the weak and strengthen the rule with advantageous timing, but may lose power otherwise. We show that the probability of successful expropriations has important implication in how the elite

[9]This is consistent with the literature of agenda setting. McCombs and Shaw (1972), followed by Scheufele (1999), Scheufele and Tewksbury (2007), and recently Helfer (2016) discuss the role of media, in determining whether the relevant parties are aware of the potential changes of other members, or the potential policies that are put into motion. In the context of Kazakhstan and Kyrgyzstan, Schatz (2009) shows the authoritarian elites make effective use of agenda setting tools to strengthen the rule.

determines the economic development policies in Section 5.1.

The first-mover advantages in mobilizing allies during regime changes are consistent with reality. In the attempted coup d'état against Recep Erdoğan of Turkey in July 2016, the failed efforts of the rebellions to block social media and the swift reaction of Erdoğan[10] jointly determined the failure of the coup. As a comparison, the 1969 Libyan coup succeeded because the coup soldiers cut off the connection between King Idris and his potential supporters, which made King Idris the last mover in power struggles.

3.2 Elite persistence

In this section we analyze how 3-agent equilibrium structures respond to power perturbations. Power perturbations are small positive or negative shocks on individual agents. We show that, regardless of the equilibrium structure and the direction of the perturbation, power in the post-perturbation structure always becomes weakly more concentrated.

When the equilibrium structure is a dictatorship and the elite agent is perturbed, either $(0, 0, c+\epsilon)$ or $(0, 0, c-\epsilon)$ remains an equilibrium and remains a dictatorship. When the equilibrium is a duopoly, a perturbation makes the slightly stronger agent dictate, i.e. $(a+\epsilon, a, 0) \to (a+\epsilon, 0, 0)$ or $(a-\epsilon, a, 0) \to (0, a, 0)$.

When the equilibrium structure is an oligarchy, $(c-b, b, c)$, power is balanced between the elite agent and the non-elites. We separate the perturbations into the ones that increases equality within the coalition, and the ones that reduce equality. For perturbations that increase equality, suppose a small positive shock $\epsilon > 0$ hits the weak, creating new power structure $(c-b+\epsilon, b, c)$, or a small negative shock hits the elite, creating new power structure $(c-b, b, c-\epsilon)$. In both cases, the elite agent receives lower rents than before. Thus the stable and rent-improving alliance $\left(\gamma_m^{'}, \gamma_e^{'}\right) = (b, b)$ becomes more attractive to the elite. Applying Proposition 1, the post perturbation equilibrium structures are $(c-b+\epsilon, b, c) \to (c-b, b, c)$ and $(c-b, b, c-\epsilon) \to (c-b-\epsilon, b, c-\epsilon)$, respectively. In both cases, the elite agent restores rents of $\frac{1}{2}$. Next, for perturbations that reduces equality, such as $(c-b-\epsilon, b, c)$ or $(c-b, b, c+\epsilon)$, by Proposition 1 the structures directly collapse to dictatorships: $(c-b-\epsilon, b, c) \to (0, 0, c)$ and $(c-b, b, c+\epsilon) \to (0, 0, c+\epsilon)$. To summarize, small changes in favor of equality are usually ineffective in oligarchies, as long as the benefits

[10] One hour after the coup started, rebellion soldiers occupied Taksim Square in central Istanbul and then inside the buildings of the state broadcaster, the Turkish Radio and Television Corporation (TRT), in Ankara, trying to control the media. They also tried to block social media such as Twitter, Facebook and Youtube. At the time, Erdoğan was on holiday outside Turkey. Two hours after the coup, Erdoğan did a Facetime interview with CNN Türk, in which he called upon his supporters to fight against the coup soldiers. Erdoğan returned to Istanbul five hours after the coup, and later wiped out the coup soldiers with the support of Turkish military.

of the vested interest group need to be satisfied. However, small changes in favor of inequality might lead to drastic descent into dictatorship because an unchecked power is created. In either case, the prestige of the elite persists, if not strengthens.

In reality, the instances where a weakened elite regains power are not uncommon. In authoritarian regimes, transitions of leadership are often from experienced old ruler to less powerful successors; for example, Kim Jung-il to Kim Jung-un of North Korea, and Hafez to Bashar al-Assad in Syria. Both sons were less experienced in politics and were not raised as successors to the throne for the majority of their youth. However, both managed to seize power while the old lieutenants ceded theirs. Similar persistence of elite families and identities are documented in other countries[11]. In the next section we discuss, in more detail, why and how the persistence can be sustained in a general coalition with more than three agents.

4 Equilibrium Characterization and Elite Persistence

In Section 3 we characterized 3-agent equilibrium structures and showed that power concentrates to the elite regardless of the direction of perturbations in 3-agent coalitions. Therefore, the position of the elite persists after perturbations. In this section, we characterize the equilibrium structures for general coalitions, and then we provide a tight condition about when such elite persistence extends to environments with more than three agents.

4.1 Equilibrium characterization

In general, agents seek alliances to eliminate others and split resources within a smaller circle. Using Proposition 1, we follow an induction approach. Suppose we have solved all $(n-1)$-agent equilibrium structures under all agendas ($n \geq 4$), we now turn to solve the equilibrium for a n-agent coalition under a specific agenda s. As in 3-agent coalitions, a sub-group of agents in a n-agent coalition choose to form a smaller coalition among themselves if they are able to adjust into a power structure which is strong enough to win in the current power struggles, is stable in all future struggles, and raises rents for participating agents. Formally, we define power structures satisfying such properties to be *stable improving*:

Definition 1. A sub-coalition $C \subset N$ with power structure γ_C is *stable improving* given (N, γ, s) if there exists an adjustable structure $\gamma_C' \leq \gamma_C$, such that: $\sum \gamma_{i \in C}' > \sum \gamma_{j \in N \setminus C}$,

[11] See Stone (1990) and Paige (1997) for Central American countries; Dalton (1968) for Liberia; Woodward (1955) and Wiener (1982) for US South after the civil war.

, and $U_i(C, \gamma_C^*) \geq U_i(N, \gamma)$ for all $i \in C$, where $\gamma_C^* = \phi\left(\gamma_C^{'}|s\right)$. In particular, (C, γ_C) is *strictly stable improving*, if $U_i(C, \gamma_C^*) > U_i(N, \gamma)$ for all $i \in C$.

Moreover, such smaller coalition should not be blocked by the mimicking behavior of a prior-moving agent according to the agenda. Denote $N_{\succ_s i}$ as the set of agents who move before agent i in agenda s. The existence of stable improving sub-coalitions with no powerful prior-moving agents means some agents are able and willing to form their own alliances and expel the others. In fact, the equilibrium structure can be bounded by the non-existence of strictly stable improving sub-coalitions and the non-existence of an early-moving blocking agent in such sub-coalitions.

Theorem 2. *For $|N| \leq 9$, a coalition is in equilibrium if and only if there does not exist a sub-coalition C that is strictly stable improving.*

For $|N| \geq 10$, a coalition is in equilibrium if there does not exist a sub-coalition C that is strictly stable improving, and only if there does not exist a sub-coalition C that is strictly stable improving and $\gamma_{i\in C}^{'} > \max\{N_{\succ_s i} \backslash C\}$ for all $i \in C$.

Proof. See Appendix. □

A summary of the proof follows. If a coalition (N, γ) starts in equilibrium with a stable improving sub-coalition, C, then agents in such sub-coalition should burn down to $\gamma_C^{'}$ to raise their rents. They are able to do so because there is no powerful early-mover who can block the burning process by mimicking ($\gamma_{i\in C}^{'} > \max\{N_{\succ_s i} \backslash C\}$). Therefore, the original power structure cannot be in equilibrium. On the other hand, if a power structure does not have a stable improving sub-coalition, then for every adjustable winning sub-coalition, C, such that $\gamma_C^{'} \leq \gamma_C$ and $\sum \gamma_{i\in C}^{'} > \sum \gamma_{i\in N\backslash C}$, there is some agent who enjoys no more rents in the deviation than in the current power structure. Consequently, the deviation proposal fails to pass and the original structure is in equilibrium. However, the theorem does not eliminate the possibility of an equilibrium structure where a rent-improving regime change is blocked by a powerful agent who has first-mover advantage, and enjoys more rents in the status quo.

Theorem 2 establishes a complete characterization for small-sized coalitions, and a necessary condition and a sufficient condition to bound the candidates of equilibrium power structures for large-sized coalitions under a given agenda. More importantly, it characterizes the two ways to support an equilibrium structure in general. The first way is the non-existence of strictly stable improving coalitions. It frees the agents from the concern of "my friend today becomes my foe tomorrow". The concept resembles that of AES in that a power structure is stable when there does not exists a stable alternative. There are however, two major differences: 1. AES does not allow deviations after deviations because power levels are fixed; 2. The algorithm in AES is independent from agenda, while in our algorithm different agendas leads to different equilibrium structures. The second way to support an equilibrium structure is to have one (or more) powerful early-moving agent, or a *Steward*, who enjoys staying in the current power structure and is able to block a strictly stable improving coalition by threatening to mimic their burning choices.

Notice that in the definition of stable improving coalitions, the adjustable structure $\gamma_C^{'}$ might not coincide with equilibrium structure γ_C^*. That is, agents may aim for temporarily unstable deviation, if the stable deviation after the deviation is desirable. This echoes the idea of far-sighted stability in Chwe (1994). It is crucial when in some power structures, $\gamma_C^{'}$, is winning while γ_C^* is not ($\sum \gamma_{i\in C}^{'} > \sum \gamma_{i\in N\backslash C} \geq \sum \gamma_{i\in C}^*$), that agents in C are willing to deviate to $\gamma_C^{'}$ to eliminate agents in $N\backslash C$ and further adjust to γ_C^*.[12] Therefore, the non-existence of strictly stable improving coalitions is stricter than rent-improving for deviating agents because it requires the deviation to be in equilibrium as well. But it is weaker than coalition-proof equilibrium (Bernheim et. al 1987) because it allows for deviations after deviation.

4.2 Elite persistence in general

In Section 3, the reason for the persistence of the elite in an oligarchy is that the inner circle consisting of the medium agent and the elite remains intact during equality-enhancing perturbations. Moreover, power struggles triggered by perturbations only stop when the outsiders cede just enough power to accommodate the insiders. In this section, we show that such intuitions carry to a general environment where power burning actually takes place during the formation of equilibrium.

When an agent burns power for survival, there are two cases. In the first case, he faces inevitable elimination thus has to burn all of his power and give away the rights to claim rents. For example, in the power structure $(1, 2, 6, 8)$, the agent with power 1 can avoid elimination only by burning power to 0. In the second case, he burns power partially to block a stable improving sub-coalition. As in Section 3, the agent with power 2 in $(2, 4, 5)$ burns power partially to block the stable improving coalition of $(4, 4)$, which creates the equilibrium structure $(1, 4, 5)$. For the block to succeed, some agents in the original stable improving coalition should be indifferent between the rents received in the deviation and the

[12] An example is $\gamma = (3, 4, 6, 6, 6)$, when the agenda is bottom-up, i.e. from the weakest agent to the strongest. A stable improving sub-coalition C has power structure $\gamma^{'} = \left(\gamma_1^{'} = 3, \gamma_2^{'} = 4, \gamma_3^{'} = 6\right)$, and $\gamma_C^* = (\gamma_1^* = 2, \gamma_2^* = 4, \gamma_3^* = 6)$. Anticipating this, agent 4 and 5 have to burn down to avoid elimination, and the equilibrium is $(2, 4, 6, 0, 0, 0)$. In equilibrium, the first agent burns power $(3 \to 2)$, but enjoys higher rents $\left(\frac{2}{12} > \frac{3}{25}\right)$.

post-burning structure. In general, for a coalition N, with initial structure γ and equilibrium structure γ^*, denote N_+ as the coalition of agents with strictly positive power. We have the following lemma.

Lemma 1. *When power is partially burnt during power struggles, some agent in the equilibrium structure receives the same rents in the current structure as in a stable improving alternative. That is, if* $0 < \gamma_i^* < \gamma_i$*, for some* $i \in N$*, then there exists* $C \subsetneqq N_+$ *with* $\gamma_C^{'}$*, where* $\gamma_C^{'} \leq \gamma_C^*$ *and* $\gamma_C^{'} > \gamma_{N \setminus C}^*$*, such that for some* $j \in C$*,* $U_j(N, \gamma^*) = U_j\left(C, \phi\left(\gamma_C^{'}\right)\right)$

Proof. See Appendix. □

Lemma 1 captures a key economic force within the model. Agents, facing the threat of elimination, give away power to the extent that the deviating agents are indifferent between the rents from current structure versus from the stable improving sub-coalition. Denote the set of agent(s) whose deviation constraint binds as G, $G = \left\{x \in C | U_x(N, \gamma^*) = U_x\left(C, \phi\left(\gamma_C^{'}\right)\right)\right\}$. We call the agents in G "Guardians" because they are the elites who guard the stable structure from deviations. Denote the complement set of C in N as $A = N \setminus C$. We call agents in A "Accommodators" because they give away power to accommodate the guardians. Applying this terminology to the example in Section 1, where $\gamma^* = (\gamma_w^* = 1, \gamma_m^* = 4, \gamma_e^* = 5)$. The stable improving coalition with power structure $\gamma_C^{'} = \left(\gamma_m^{'} = 4, \gamma_e^{'} = 4\right)$ is blocked by the accommodator, w, who partially burns $\gamma_w^* = 1 < \gamma_w = 2$, to accommodate the guardian e, who enjoys the same rents in the equilibrium $\left(\frac{5}{1+4+5}\right)$ as in the deviation $\left(\frac{4}{4+4}\right)$. In Michels' (1915) words, the guardians are the elites that " (the) society cannot exist without a dominant or political class, and that the ruling class, while its elements are subject to frequent partial renewal, nevertheless constitutes the only factor of sufficiently durable efficacy in the history of human development."

As with the 3-agent coalitions, we are interested in whether shocks that weaken the strong (the guardians) and shocks that empowers the weak (the accommodators) actually lead to increase of equality in equilibrium. In both cases, we focus on equilibrium structures where partial power is burnt during the formation of equilibrium; that is, the set of guardians is non-empty.

Suppose a small negative shock hits one of the guardians. There are two possible outcomes. First, if there are multiple guardians and one of them is weakened, then the rest of the guardians now enjoy even higher rents than in the stable improving alternative. Consequently, the deviation remains unattractive to them and no power struggles take place. Second, suppose there is only one guardian who is hit by a small, negative shock. The weakened guardian receives fewer rents, thus allowing for a *strictly* stable improving. As a result, the accommodators respond by burning their own power to appease the guardian. Formally, we have the following theorem.

Theorem 3. *If $0 < \gamma_i^* < \gamma_i$ for some $i \in N$, and G is a singleton, then there exists a threshold, $\underline{\eta} > 0$, such that for any negative perturbation, $0 < \eta < \underline{\eta}$, on $x \in G$, $U_x(\gamma^*) = U_x\left(\phi\left[\left(\gamma_x^* - \eta, \gamma_{-x}^*\right)\right]\right)$.*

Proof. See Appendix. □

Theorem 3 illustrates the persistence of the elite after they are weakened. The elite recovers from negative shocks, in terms of the rents received, because the stable improving coalition remains intact after the shock. Thus, the "outside option" of the elite remains unchanged, and the accommodators bear the burden of the negative shock upon the power of the elite. In Michels' (1915) words, "The government, or, ... the state, cannot be anything other than the organization of a minority. It is the aim of this minority to impose upon the rest of society a legal order which is the outcome of the exigencies of dominion and of the exploitation of the mass." Such idea, known as "The Iron Law of Oligarchy", is well documented in the studies of power structures. In fact, Kendall-Taylor and Frantz (2016) studied all 79 dictatorships between 1946 and 2012, and concluded, "...a leader's death in office almost never leads to significant near-term liberalization. Likewise, only rarely does it spell the end of the regime or precipitate instability in the form of coups or protests. On the contrary, authoritarian regimes have proven to be remarkably resilient when a leader dies." In the meantime, the trend of organizations towards oligarchy continued to spread throughout the first decade of the twenty-first century. According to Magaloni and Kricheli (2010), one-party regimes have now become the most common type of authoritarian rule, constituting 57% of the authoritarian regimes during 1950–2006 and 33% of the total number of regimes in the world.

Meanwhile, if a positive shock hits the guardian, he always receives at least as many rents as before, because the guardian always has the option to burn up the positive shock and returns to the previous equilibrium structure. Therefore, an immediate corollary follows from Theorem 3 is that, regardless of the directions of perturbations, the power of the strong always gets weakly more centralized.

Corollary 2. *If $0 < \gamma_i^* < \gamma_i$ for some $i \in N$, and G is a singleton, then there exists $\underline{\eta} > 0$, such that $U_x(\gamma^*) \leq U_x\left(\phi\left[\left(\gamma_x^* + \eta, \gamma_{-x}^*\right)\right]\right)$, for any $\eta \in \left(-\underline{\eta}, \underline{\eta}\right)$ applied on $x \in G$.*

Next, we consider the other side of the same coin: the shock that increases the power of an accommodator. When an accommodator is empowered, the guardian is necessarily weakened, which again triggers a strictly improving sub-coalition.

Theorem 4. *If* $0 < \gamma_i^* < \gamma_i$ *for some* $i \in N$*, then there exists* $\overline{\eta} > 0$ *such that* $U_x(\gamma^*) = U_x\left(\phi\left[\left(\gamma_j^* + \eta, \gamma_{-j}^*\right)\right]\right)$*, for any* $0 < \eta < \overline{\eta}$ *and* $\gamma_j^* < \gamma_j$*,* $j \in A$*.*

Proof. See Appendix. □

Theorem 4 illustrates that the policy of empowering the weak to increase social equality might not work, because the stable improving coalition involving the elites always needs to be accommodated. To an extreme it can be totally ineffective (See Section 3). This theorem provides a unified explanation of the consolidation of autocratic regimes facing opportunities in favor of the non-elites. Such opportunities may be trade, as in Guatemala and Mexico in the 19th century; or the abolition of slavery as in West Africa (Robinson 2012); or the abolition of Mita in Peru (Dell 2008). Therefore, one should not take the rise of the non-elites for granted when a favorable junction arrives. In fact, as long as the elite circle remains intact, empowerment of the weak seldom works.

Additional remarks. Sometimes social empowerment does induce dramatic changes of the political landscapes. Such empowerment usually involves significant increase of power, or changing both power *and* agenda at the same time. In the Arab Spring, the Tunisian government under Ben Ali, focused on foreign direct investment, open trades, and open media, which not only empowered the masses but also changed domestic political agendas. The success of the revolution reveals the possible coordination between social empowerment and agenda-settings.

5 Applications

In this section, we explore the implications of adding elements of real life to the framework. In Section 5.1, we introduce the possibility of production, and discuss the boundary of power struggles under the tradeoff between expropriation and economic development. In Section 5.2, we consider the impact of a foreign force on domestic power structures. Throughout Section 5, we assume that initiating power struggles is a *choice* for some particular agent in the coalition[13], and we examine how that choice interacts in the presence of other forces. We also aim to generate testable predictions, in order to demonstrate that the framework goes beyond a technical practice.

[13] In reality, initiation of power struggles can indeed be a policy choice. The national development of China was guided by the principle of "taking class struggles as central tasks" since 1949, especially during the Cultural Revolution, before the ruling party decided to "concentrate on modernization" in 1977. Vietnam turned to "Renovation" in 1986 with the goal of creating a "socialist-oriented market economy", after 15 years of leftist policies.

5.1 Expropriation vs. development

Power struggles aim to change how the pie is split, whereas production aims to enlarge the pie. If the choices are mutually exclusive, a natural trade-off exists. From a social welfare perspective, production is always preferred to power struggles. Intuitively, a ruler prioritizes economic development when he is free from survival concerns (Gehlbach and Keefer 2011). In the section, we show that this intuition may fail. A ruler free from survival concerns may focus on expropriations, while a ruler with survival concerns may focus on economic development.

Players. Consider a society consisting of three representative agents: the elite agent (e) who rules; the magistrate (m) who helps to rule; and the worker (w) who produces. Denote the initial power structure $\gamma_0 = (\gamma_w^0 = k + h, \gamma_m^0 = 1 - k, \gamma_e^0 = 1)$, where $k > 0$, $h \geq 0$. By Proposition 1, the structure is in equilibrium when $h = 0$, thus h measures how decentralized power is from the equilibrium. We also assume that $k + h < 1 - k$, such that the magistrate holds more power than the worker.

Strategies. The game begins with the initial power structure γ_0 and agenda s. The elite chooses whether to initiate power struggles. If power struggles are initiated, the game is played as in Section 2, and agents receive rents according to the equilibrium structure. If the ruler chooses not to initiate power struggles, then the worker produces. Production increases the worker's power by $\beta > 0$. And production plus the rents are distributed according to $\gamma_0^{'} = (k + h + \beta, 1 - k, 1)$.

Payoffs. There is an existing pie of size π to share. Production enlarges the pie by $\alpha\gamma_w^0$, where $\alpha > 0$ is a measure of productivity.

An intuitive approach suggests a direct comparison between the increased rents from expropriation (power struggles) and the increased size of the pie, due to production. However, as Proposition 1 shows, the elite, in a 3-agent coalition, may lose power in the struggles, which makes expropriation risky. We therefore distinguish elites by their likelihood of survival in power struggles, and provide their optimal decisions in the following proposition:

Proposition 2. *Rulers with survival concerns prioritize economic growth over expropriation more often than those without survival concerns. That is:*

When $s \neq (m, w, e)$ and $s \neq (w, m, e)$, the ruler focuses on production if and only if

$$\frac{1}{2}\pi\beta + \left(\frac{1}{2}\pi - \alpha\right) \leq k < \frac{1-h}{2}$$

.

When $s = (m, w, e)$ or $s = (w, m, e)$, the ruler focuses on production if and only if

$$\min\left\{\frac{1}{2}-h,\ \frac{1}{2}\pi\beta+\left(\frac{1}{2}\pi-\alpha\right)\right\}\le k<\frac{1-h}{2}$$

.

Proof. Without concerns of losing power, the ruler conducts a direct comparison between enlarging the pie and getting a larger share of the existing pie: $\frac{1}{2+h+\beta}\left[\alpha\left(k+h\right)+\pi\right]>\frac{1}{2}\pi$, which gives the left hand side of the first inequality, $k>\frac{1}{2}\pi\beta+\left(\frac{1}{2}\pi-\alpha\right)h$. The right hand side of the first inequality is given by the assumption that $k+h<1-k$.

When the ruler has survival concerns, he decentralizes whenever the worker is strong enough to overthrow him in agenda $s=(m,w,e)$ or $s=(w,m,e)$, therefore he decentralizes whenever $k+h>\frac{1}{2}$. When $k+h\le\frac{1}{2}$, the elite ruler has no survival concerns and behaves the same as in the first case. Combining the two we get the expression on the left hand side of the second inequality. □

It is intuitive that power struggles are more likely to be reduced by economic development: when productivity (α) is higher, when the existing pie is smaller, and when the empowerment from production is insignificant (β is low). As a result, high value-added agents are less likely to be expropriated[14], whereas a regime abundant in natural resources often suffers from power struggles.

The key implication of the proposition is that elite rulers, facing survival concerns, favor economic development in a larger set of initial power structures. This is because an elite ruler is forced to prioritize economic development when expropriation would trigger a revolution to overthrow the ruler.

Prediction: Timing of economic reform. The analysis above generates testable predictions of when a regime turns from "extractive" to "inclusive", per Acemoglu and Robinson (2013) terminology. When the current power structure is highly skewed, elite rulers behave in the same manner under all agendas: promoting economic growth when productivity is high, and extracting rents from others when existing resources are abundant. When the current structure is less skewed, elite rulers, under the agendas where they have survival concerns, tend to prioritize economic development more often than others. On the other hand, we predict a strong ruler follows a "sow and reap" strategy in a dynamic environment. He should aim for economic development for a number of periods, then initiate power struggles before the potential opponents become too powerful, and, finally, resume economic development thereafter. This strategy could explain the variations, and sometimes the cyclicality, of development policies seen in a regime across time or across different regimes.

[14] For the opposite direction, a Chinese proverb best describes the ruler's strategy: after the cunning hare is killed, the hound is boiled.

5.2 Domestic vs. international conflicts

In the previous discussions, power struggles occur domestically; i.e., the grand coalition is fixed. In this section, we consider how domestic power struggles are affected by the presence of external power. It is convenient to think of power struggles as conflicts such as wars or negotiations over resources. Intuitively, imminent foreign threats soften domestic conflicts, because a united domestic power has a better chance against the threats. On the other hand, an agent to be expelled can make himself indispensable in the power balance by inviting a strong foreign force to settle the disputes. We thus argue that the both external threats and international cooperation may help alleviate domestic power struggles.

5.2.1 Foreign Threat

Players. Considers a domestic country with three warlords, w, m and e. The current power structure is $\gamma_0 = (\gamma_w = k + h, \gamma_m = 1 - k, \gamma_e = 1)$. The agenda is top-down. A foreign country with power level g, $2 < g < 2 + h$, attacks the domestic country with probability $0 \leq \mu \leq 1$. The war is won by domestic forces if and only if they are collectively more powerful than the enemy $\gamma_p + \gamma_m + \gamma_e \geq g$.
Strategies. The game proceeds as follows: 1. Domestic agents conduct power struggles as in Section 2; 2. The foreign country attacks; 3. The resource is distributed to the winner(s).
Payoffs. Winners of the war share the resources according to their power levels, and the loser gets nothing.

It is easy to see that a foreign threat softens domestic power struggles because $g > 2 > 2 - 2k$, thus warlords m and e are no longer interested in seeking the power burning alliance $(1 - k, 1 - k)$ because such an alliance fails to defeat the invader. Consequently, the domestic agents set aside power struggles and join hands when foreign forces are likely to invade.

Proposition 3. *Power struggles are softened when external threat is high. In particular, power struggles are ceased when $\mu > \underline{\mu}$, where $\underline{\mu} = \frac{h}{2+h}$.*

Proof. The cut-off is determined by $\left(1 - \underline{\mu}\right) \frac{1}{2} = \frac{1}{2+h}$, which gives $\underline{\mu} = \frac{h}{2+h}$. □

There are two implications of this proposition. First, the cut-off is independent of the level of external threat g, because should the invasion takes place, the outcome is always binary. Therefore, when the threat is credible ($2 < g < 2 + h$), domestic conflict resolution depends on the probability of invasion (μ), instead of th strength of the invader[15]. Second, the cut-off increases in h, because when the invasion does not take place, a resource-rich

[15] We leave an alternative model, with probabilistic victory, for future research.

warlord further incentivizes more extraction.

Prediction: Post-war struggles. We predict that, in general, power struggles are more likely to take place shortly after the end of conflicts and less likely shortly before the start of conflicts. Below we provide historical evidence of this prediction.

Georgy Zhukov was one of the most important military commander of the Soviet Union in World War II. He led and won multiple battles, including the Battle of Berlin, which put an end to the War in Europe. During the war, Zhukov held several prestigious positions in the army and in the political system because of his outstanding military talents. He served as Chief of the General Staff and then Deputy Commander-in-Chief. However, after the War, Zhukov was viewed by Stalin as a potential threat to his leadership because of his popularity in the army. Stalin replaced Zhukov with Vasily Sokolovsky as Commander-in-Chief of the Soviet Ground Forces. Zhukov was assigned to command the Odessa Military District, away from Moscow and lacking in strategic significance and troops. Zhukov's military career during the 1940s illustrates how power struggles respond to external threats. Power struggles can be softened when the threat is imminent, and will resume after the threat is exterminated.

Soviet Union is not a lone example. In China, during World War II, Kuomintan and the Communist Party of China, who were previously in civil war, united forces against Japan. The two parties resumed the civil war after defeating the Japanese. Moreover, in the 7th century, the Spanish kingdoms ceased the conflict to fight against the Moors.

5.2.2 Foreign Ally

Players. Consider a similar domestic country with three political parties, representing the workers, the middle class, and the elite with initial domestic structure $\gamma_0 = (\gamma_w = k+h, \gamma_m = 1-k, \gamma_e = 1)$. The agenda is (m, w, e). The elite has connections with a foreign ally (a), who holds power $1 < d < 2$. For simplicity, the foreign ally is assumed to move after domestic players, if it is invited to participate in domestic affairs. Thus, the revised agenda is (m, w, e, a).

Strategies. The game proceeds as follows: 1. The elite decides whether to invite a foreign ally; 2. Power struggles take place as in Section 2, either in a 3-agent or a 4-agent coalition; 3. Payoffs are assigned.

Payoffs. Agents split rents of 1 according to their power levels after the power struggles.

By Proposition 1, when $k + h > \frac{1}{2}$, the worker and the middle class form an alliance to overthrow the elite in a 3-agent coalition. Anticipating the elimination, the elite has incentive to invite the foreign ally, which transforms the power structure to $(d + k - 2, 1 - k, 1, d)$. This is the equilibrium of the 4-agent coalition after power struggles. The elite receives

$\frac{1}{2d} > 0$, which is preferred to ceding power completely, as they would in the 3-agent coalition.

In general we have the following result for any coalition with domestic conflict, $\phi(\gamma) \neq \gamma$.

Proposition 4. *Any domestic conflict can be stopped by a foreign power. That is, for any coalition N with domestic conflicts such that $\phi(\gamma|s) \neq \gamma$, there exists an external agent with power level γ_f such that $\phi((\gamma, \gamma_f)|s) = (\gamma, \gamma_f)$.*

Proof. Take $\gamma_f = \sum \gamma_{i \in N}$, by Theorem 2, $\phi((\gamma, \gamma_f)) = (\gamma, \gamma_f)$ □

Notice that despite the disappearance of conflict, this might not always be in the interest for the elite to do so.

Prediction: International interventions. We predict that the local elites may cede power to a foreign superior power when their domestic rule is threatened. In addition, we predict that the prestige of the local elites may persist among local communities after inviting a foreign power to join domestic affairs. Below we provide evidences from the 19th century Caribbean islands and modern-day Sri Lanka.

In the 19th century, there were fourteen Caribbean, British sugar colonies[16]. Here the sugar cane plantation owners were the dominant social class, and slaves were the major source of the labor supply. In 1833, the British parliament passed *An Act for the Abolition of Slavery*, which emancipated the slaves. Consequently, the poor were significantly empowered and began to play a larger role in local governance. According to Carvalho and Dippel (2016), "Starting with Montserrat in 1861 and ending with Grenada in 1876, all but one of the powerful assemblies voluntarily dissolved themselves, and invited the Crown to write a new constitution for them with a legislature appointed by the governor". Ashdown (1979) argues that "the colonies gave up their elected assemblies voluntarily, for in most cases the white, privileged classes preferred direct imperial government to the government of the colored classes who were slowly obtaining greater representation in the legislative councils."

Conversely, if the current domestic conflict cannot be resolved by domestic powers alone, a local force may invite external allies to eliminate the threats for them. The civil war in Sri Lanka had lasted for more than 25 years, and was ended with the external help from China, who "provided lethal military equipment and post-war diplomatic support" [17]). In exchange for their aid, Sri Lanka has signed a $1.1billion deal with China for the control and development of the southern deep-sea port of Hambantota, which has potential military

[16]Namely, Antigua, Barbados, Jamaica, Montserrat, Nevis, St. Kitts, the Virgin Islands, Dominica, Tobago, St. Vincent, Grenada, Trinidad, St. Lucia, and Guyana. For a detailed discussion of the historical background, see Carvalho and Dippel (2016).

[17]See: https://www.brookings.edu/blog/order-from-chaos/2015/05/01/india-v-china-in-sri-lanka-lessons-for-rising-powers/

uses[18]. For more examples of the power balancing role of external intervention, see Levine and Modica (2016).

6 Discussion and Conclusion

The goal of this paper is to study the persistence of elites in coalition dynamics. It does so by constructing a model of power struggles, where players use their power to eliminate others and split resources in a coalition. The model innovation is to allow players to cede power when they are about to be expelled. A power structure is in equilibrium when the opponents of the current structure fail to adjust to a stable, rent-improving alternative. We characterize stable structures in general and show that power often ends up more concentrated to a few elite members when equilibrium structures are perturbed. We use the model to explain why social and political hierarchies often prove extremely difficult to eradicate and why, in many cases, revolutions that aim to install social equality quickly reproduce the same type of hierarchies that the revolutionaries sought to destroy.

It is important to point out that, although commitment is not possible during power adjustment and power struggles, we do assume that when a burning proposal reaches unanimous agreement, the corresponding burning outcome will be implemented. Alternatively, if we assume agents can only make individual burning choices, Theorem 2 fails to hold and the set of stable structures enlarges. The intuition is that a stable improving coalition may not be implemented if other undesirable deviations are triggered during the transition. We provide a counter-example in Appendix B. A future direction of the research is to characterize stable coalitions with individual burning.

Another assumption we make in the model is costless eliminations. It is an approximation of the case where rents received by staying in the coalition significantly outweigh the potential costs of elimination. Adding a positive elimination cost does not change the resurgence of hierarchies nor the persistence of elites in perturbations. However, drastic regime changes from oligarchies to dictatorships take place less frequently with positive elimination costs. To see this, suppose every elimination costs $\kappa > 0$ to each agent involved, the equilibrium structure of a 3-agent coalition $(2, 4, 5)$ becomes $\left(1 + \frac{20\kappa}{1-2\kappa}, 4, 5\right)$ because the stable improving alternative of $(4, 4)$ only guarantees rent of $\frac{1}{2} - \kappa$ instead of $\frac{1}{2}$. As a result, if the elite agent is hit by a small positive perturbation, $\left(1 + \frac{20\kappa}{1-2\kappa}, 4, 5 + \epsilon\right)$, the perturbed structure remains in equilibrium since $1 + \frac{20\kappa}{1-2\kappa} + 4 > 5 + \epsilon$. That is, positive elimination costs offer equilibrium structures some slackness in how they can perturbed[19].

[18]See: http://www.bbc.com/news/world-asia-40761732

[19]Of course, costs of elimination may depend on the power difference of involving agents, and may be

A key difference between our model and AES is that we allow for fully flexible power adjustment while AES assumes fixed power levels[20]. The real world lies somewhere between the two. One way to reconcile the two models is to introduce a lower bound of power burning: individuals can adjust their power only above a lower bound. The lower bound may be determined by blood lines, religious beliefs, ethnicity (Montalvo and Reynal-Querol 2005), institutional legacies (Evans 1995), etc. For example, if one has a royal name or is of royal blood, it is often not enough for him or her to cede power and claim loyalty to the current ruler. The leader sometimes ends up eliminating the bloodline to tie up the loose ends. This practice endures from the sibling killing tradition in Ottoman Empire to the assassination of Kim Jung-Nam in February, 2017, and the recent arrest of Prince Alwaleed bin Talal and Prince Mutaib bin Abdullah of Saudi Arabia.

Lastly, we discuss several other directions we would like to explore in the future and their respective intuitions. First, the current analysis focuses on non-targeted transfers (burning)[21]. A direct generalization would be to allow targeted transfers, where a member can transfer power precisely to another individual. We conjecture that the qualitative results of the model will not change, because allowing targeted transfers provides a cheaper way to accommodate pivotal players. For example, if free transfers are allowed and we start from a 3-agent coalition $(3, 4, 5)$ under a top-down agenda, the cheapest way for agent 1 to accommodate agent 3 is to directly transfer 1 piece of power; that is, to turn the power structure to $(2, 4, 6)$. It differs from the original equilibrium structure $(1, 4, 5)$, but they share the feature of being oligarchies.

Second, the coalition in the current framework is fully connected: every agent can form alliances with everyone else. It would be interesting to investigate how particular network structure affects coalition formation outcomes for the following reasons: First, restrictions on coalition structures are pervasive in reality. Members in a large society may not know

asymmetric even for agents with same power, due to, for instance, geographical reasons. Caselli et. al (2015) showed that conflicts are more likely when the resources of interest located near the border of two countries because it is less costly for the attacking side to occupy the resources.

[20]It is true that sometimes power struggles involve executing your enemies such as in the Great Purge in the Soviet Union. But in many other cases the struggles end with the relevant parties ceding power but keeping their lives. Even in Soviet Union, the power transition since Nikita Khrushchev, despite the struggles, had been largely without blood shed. The same applies to the transition after Cultural Revolution in China, the transition from Chiang Ching-kuo to Lee Teng-hui in Taiwan, the transition from Thein Sein to Aung San Suu Kyi in Myanmar, and the power consolidation of Bashar al Alssad in Syria.

[21]Burning $\Delta = \gamma_i - \rho_i(\gamma_i)$ pieces of power for agent i is equivalent to cede power $\frac{\Delta}{n-1}$ uniformly to every other agent in the coalition. Burning power is a special case of non-targeted power transfer. We adopt power burning as the only way of power transfer to emphasize the difficulty of such process. Unlike physical assets, power is often hard to transfer from one to another, for instance, for an army to switch allegiance to another commander. Other examples of burning power include resigning or exiting from office (Steve Bannon), migrating to a foreign country (Leon Trotsky), abandoning access to weapons (chemical, nuclear and other weapons of mass destruction), etc.

to each other or the impact of power burning is only restricted within a sub-group. In the meantime, power struggles, or conflicts of wars can be means to access to a larger coalition, as described in Dziubinski et. al (2017). Second, particular network structures speak to different types of collective action problems. For instance, if a sub-coalition has a "line" structure (eg. a command chain), the sequence of moves is in the coalition. Meanwhile, in a "star" structure (eg. a radio station with its audience), only the first mover is clear, but not the rest of the sequence.

References

[1] Daron Acemoglu, Georgy Egorov, and Konstantin Sonin. Coalition formation in political games. Technical report, National Bureau of Economic Research, 2006.

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

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

[4] Daron Acemoglu and James A Robinson. *Economic origins of dictatorship and democracy*. Cambridge University Press, 2005.

[5] Daron Acemoglu and James A Robinson. De facto political power and institutional persistence. *The American economic review*, 96(2):325–330, 2006.

[6] Daron Acemoglu and James A Robinson. Persistence of power, elites, and institutions. *The American Economic Review*, 98(1):267–293, 2008.

[7] Daron Acemoglu and James A Robinson. *Why nations fail: The origins of power, prosperity, and poverty*. Crown Business, 2013.

[8] Daron Acemoglu and James A Robinson. The emergence of weak, despotic and inclusive states. 2017.

[9] Philippe Aghion, Alberto Alesina, and Francesco Trebbi. Endogenous political institutions. *The Quarterly Journal of Economics*, 119(2):565–611, 2004.

[10] Philippe Aghion and Patrick Bolton. Incomplete social contracts. *Journal of the European Economic Association*, 1(1):38–67, 2003.

[11] Peter Ashdown. *Caribbean history in maps*. [Port of Spain], Trinidad: Longman Caribbean, 1979.

[12] B Douglas Bernheim, Bezalel Peleg, and Michael D Whinston. Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory*, 42(1):1–12, 1987.

[13] Jean-Paul Carvalho and Christian Dippel. Elite identity and political accountability: A tale of ten islands. Technical report, National Bureau of Economic Research, 2016.

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

[15] Michael Suk-Young Chwe. Farsighted coalitional stability. *Journal of Economic theory*, 63(2):299–325, 1994.

[16] Ernesto Dal Bó, Pedro Dal Bó, and Jason Snyder. Political dynasties. *The Review of Economic Studies*, 76(1):115–142, 2009.

[17] George Dalton. History, politics, and economic development in liberia. *The Journal of Economic History*, 25(4):569–591, 1965.

[18] Bruce Bueno De Mesquita. *The logic of political survival*. MIT press, 2005.

[19] Alexis De Tocqueville. *Ancien regime and the French revolution*. Penguin UK, 2008.

[20] Melissa Dell. The mining mita: Explaining institutional persistence. *Manuscript, Department of Economics, MIT*, 2008.

[21] William Doyle. *The Oxford history of the French revolution*. Oxford Paperbacks, 1989.

[22] Marcin Dziubinski, Sanjeev Goyal, and David EN Minarsch. The strategy of conquest. 2017.

[23] Peter B Evans. *Embedded autonomy: States and industrial transformation*. Princeton University Press, 2012.

[24] James D Fearon. Self-enforcing democracy. *The Quarterly Journal of Economics*, 126(4):1661–1708, 2011.

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

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

[27] Paul R Gregory. *Politics, Murder, and Love in Stalin's Kremlin: The Story of Nikolai Bukharin and Anna Larina*. Hoover Press, 2013.

[28] Bernardo Guimaraes and Kevin D Sheedy. Guarding the guardians. *The Economic Journal*, 2017.

[29] Vaclav Havel and Paul Wilson. The power of the powerless. *International Journal of Politics*, 15(3/4):23–96, 1985.

[30] Luzia Helfer. Media effects on politicians: An individual-level political agenda-setting experiment. *The International Journal of Press/Politics*, 21(2):233–252, 2016.

[31] Thomas Hobbes. *Leviathan*. A&C Black, 2006.

[32] Karl Jandoc and Ruben Juarez. Self-enforcing coalitions with power accumulation. *International Journal of Game Theory*, pages 1–29, 2013.

[33] Andrea Kendall-Taylor and Erica Frantz. When dictators die. *Journal of Democracy*, 27(4):159–171, 2016.

[34] DM Kilgour. Equilibrium points of infinite sequential truels. *International Journal of Game Theory*, 6(3):167–180, 1977.

[35] David K Levine, Salvatore Modica, et al. An evolutionary model of intervention and peace. Technical report, UCLA Department of Economics, 2016.

[36] Beatriz Magaloni and Ruth Kricheli. Political order and one-party rule. *Annual Review of Political Science*, 13:123–143, 2010.

[37] Maxwell E McCombs and Donald L Shaw. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972.

[38] Robert K Merton et al. The matthew effect in science. *Science*, 159(3810):56–63, 1968.

[39] Robert Michels. *Political parties: A sociological study of the oligarchical tendencies of modern democracy*. Hearst's International Library Company, 1915.

[40] Mancur Olson. *Power and prosperity: Outgrowing communist and capitalist dictatorships: Outgrowing communist and capitalist dictatorships*. Basic books, 2000.

[41] Jeffrey M Paige. Coffee and power. *Revolution and the Rise of Democracy in Central America*, 1997.

[42] Ann Paludan. Chronicle ofthe chinese emperors: The reign-by-reign record of the rulers of imperial china, 1998.

[43] Richard Pipes. *Russia under the Bolshevik regime*. Vintage, 2011.

[44] Marek Pycia. Stability and preference alignment in matching and coalition formation. *Econometrica*, 80(1):323–362, 2012.

[45] Debraj Ray and Rajiv Vohra. The farsighted stable set. *Econometrica*, 83(3):977–1011, 2015.

[46] James A Robinson. Elites and institutional persistence. *The role of elites in economic development*, pages 29–52, 2012.

[47] Edward Schatz. The soft authoritarian tool kit: Agenda-setting power in kazakhstan and kyrgyzstan. *Comparative Politics*, 41(2):203–222, 2009.

[48] Dietram A Scheufele. Framing as a theory of media effects. *Journal of communication*, 49(1):103–122, 1999.

[49] Dietram A Scheufele and David Tewksbury. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication*, 57(1):9–20, 2007.

[50] Samuel Z Stone. *The heritage of the conquistadors: Ruling classes in Central America from the conquest to the Sandinistas*. Univ of Nebraska Pr, 1990.

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

[52] Li Tao. Xu zizhi tongjian changbian. *Collected data for a continuation of the Comprehensive mirror for aid in government*, 2004.

[53] Jonathan M Wiener. Social origins of the new south: Alabama, 1860-1885. 1982.

[54] Comer Vann Woodward. *The strange career of Jim Crow*. Oxford University Press, USA, 1955.

[55] Yan Bing Zhang, Mei-Chen Lin, Akihiko Nonaka, and Khisu Beom. Harmony, hierarchy and conservatism: A cross-cultural comparison of confucian values in china, korea, japan, and taiwan. *Communication research reports*, 22(2):107–115, 2005.

Appendix

Appendix A: Omitted proofs

Proof of Theorem 1.

Proof. **Existence.** First, observe that the game has finite horizon. The burning meetings last for finite rounds because power is discrete, so there are at most $\frac{\prod_{i=1}^{n}\gamma_i}{\epsilon^n}$ rounds of burning meetings. The elimination meetings last for finite rounds because there are finite proposals and finite voters, and if no proposal is accepted the status quo is enforced. By Zermelo's Theorem, a finite horizon game of perfect information admits a SPE.

Uniqueness of equilibrium. Given agenda s, suppose there exists another equilibrium $\left(N^{'},\gamma^{'}\right) \neq (N^*,\gamma^*)$. There are two cases. First, $N^* \neq N^{'}$. Without loss of generality assume that $N^* \subsetneq N^{'}$. Then there exists some agent i such that $i \in N^{'}$ but $i \notin N^*$. Therefore in the first equilibrium $U_i = -n$, which cannot happen in equilibrium because agent i can simply burn power down to zero in the previous burning meeting such that $U_i = 0$, which is an improvement. Therefore (N^*,γ^*) cannot be in equilibrium. By the same logic we know that $N^* = N$, because otherwise any agent in $N \backslash N^*$ has a better response by burning down to zero.

Consequently, elimination does not take place on equilibrium path. Now suppose $N^* = N^{'}$ but $\gamma^* \neq \gamma^{'}$. Denote the first time (round) that an agent burns power differently on equilibrium path as t, the member as j, strategy as $s^*_{j|h_t}$, $s^{'}_{j|h_t}$. Therefore the member is indifferent between two burning choices, $U_j\left(s^*_{j|h_t}\right) = U_j\left(s^{'}_{j|h_t}\right)$, which implies that $\frac{\gamma^*_j}{\sum \gamma^*_i} = \frac{\gamma^{'}_j}{\sum \gamma^{'}_i}$. Furthermore, the tie-breaking rule suggests that $\gamma^*_j = \gamma^{'}_j$. So j cannot burn power differently. Neither can any other members, $\gamma^* = \gamma^{'}$. Contradiction.

So far we have shown that equilibrium exists and is unique for given N, γ, s. There might exist multiple SPE's, because in the voting stage, if a previous member included in the proposal votes no, and the next member is indifferent between yes and no because the proposal is rejected either way. □

Proof of Corollary 1.

Proof. See the first part of the "uniqueness" proof of Theorem 1. □

Proof of Proposition 1.

Proof. When the elite agent is very powerful, it naturally dictates. When the elite agent is not powerful enough, the key is whether the weak and the medium agent could form a powerful alliance against him.

When $a \leq \frac{c}{2}$, an alliance between the weak and the medium is not powerful enough. So the unique stable and rent-improving coalition remains $\left(\gamma_m^{'} = b, \gamma_e^{'} = b\right)$. The consequent equilibrium is $(c - b, b, c)$.

Now we consider the cases under $a > \frac{c}{2}$. When $s = (m, w, e)$ or (w, m, e), the weak and the medium can form the sub-coalition $\left(\gamma_w^{'} = a, \gamma_m^{'} = a\right)$ before the elite moves. Facing (a, a, c), the elite, as the last-mover, has no choice but to burn down to zero because gambling for probabilistic inclusion is never optimal. Therefore the equilibrium structure is $(a, a, 0)$. In any other agenda where the elite is not the last mover, the elite can always mimic the medium agent's burning strategy and secure rent of $\frac{1}{2}$. As a result, the weak agent should always accommodate the elite by burning down to make up the power difference between the elite and the medium agent. Therefore the equilibrium structure is $(c - b, b, c)$. □

Proof of Theorem 2.

Proof. For a general coalition N,

"If" condition.

Suppose there does not exist a stable improving sub-coalition, that is, for any winning sub-coalition $C \subset N$ such that $\gamma_C^{'} > \gamma_{N \backslash C}$, where $\gamma_i^{'} \leq \gamma_i$, and $\gamma_C^{*} = \phi\left(\gamma_C^{'}\right)$, there exists some agent $i \in C$ that $U_i\left(C, \gamma^{*}\right) \leq U_i\left(N, \gamma\right)$. If $U_i\left(C, \gamma^{*}\right) < U_i\left(N, \gamma\right)$, then agent i rejects the proposal when C is proposed in the elimination meeting. If $U_i\left(C, \gamma^{*}\right) = U_i\left(N, \gamma\right)$, should a equilibrium with inner circle C is formed, the agents have to participate in at least 3 rounds in burning meeting, which is dominated by maintaining the status quo by the tie-breaking assumptions. Therefore, agent i rejects the proposal in the elimination meeting as well. Anticipating the rejection, other members in C have no incentives to adjust power in the burning meeting, thus the status quo coalition remains unchanged. So the original structure is in equilibrium.

"Only if" condition.

Suppose a coalition N with γ is in equilibrium, and there is a sub-coalition C with adjusted structure $\gamma_C^{'}$ that is stable improving. Pick any agent in C who makes burning proposals in the burning meeting: if he proposes $\gamma_C^{'}$, it is a better response for all involving agents in C to agree on the proposal and secures a rent-improving payoff. He is able to do so because $\gamma_{i \in C}^{'} > \max\left\{N_{\succ_s i} \backslash C\right\}$ therefore no prior-moving agent can mimic the move. Thus the original structure is not equilibrium, contradiction.

Next, we show that for small sized coalitions, a power structure is in equilibrium only if there does not exist a strictly improving sub-coalition. By Proposition 1, it is easy to check that the condition is satisfied for 3-agent coalitions. For 4-agent coalitions, any strictly improving sub-coalition ensures the most powerful player in the sub-coalition at least $\frac{1}{2}$

of the rent (as in Proposition 1), therefore it cannot be blocked by an early-mover. The practice goes on for 5-, 6-,..., 9-agent coalitions. It fails for 10-agent coalitions, for example, $(5+\epsilon, 5+2\epsilon, ..., 5+8\epsilon, 9, 10)$ under the top-down agenda, because now the 8-agent equal-powered sub-coalition guarantees rent of $\frac{1}{8}$, which is less attractive for the powerful early-movers. □

Proof of Lemma 1.

Proof. From Theorem 2, since some power is burnt, then for any adjustable sub-coalition $C \subset N$ such that $\gamma_C^{'} > \gamma_{N\backslash C}^{*}$ and $\gamma_C^{'} \leq \gamma_C^{*}$, there exists some agent $j \in C$ that $U_j(N, \gamma^*) \geq U_j\left(C, \phi\left(\gamma_C^{'}\right)\right)$.

If $U_j(N, \gamma^*) = U_j\left(C, \phi\left(\gamma_C^{'}\right)\right)$ for some j, the statement is proved. If $U_j(N, \gamma^*) > U_j\left(C, \phi\left(\gamma_C^{'}\right)\right)$ for all j and all such sub-coalition C, pick agent i as the last agent who burns partial power, $0 < \gamma_i^* < \gamma_i$, and slightly increase his power in his last burning choice. Then the inequalities for all j's still hold (because i is the last agent burning partial power. A complete burner does not change the result, otherwise they do not burn completely) and agent i enjoys more rent. Thus it is a better response for i. Agent i can continue to do so until for some $j^{'}$, $U_{j^{'}}(N, \gamma^*) = U_{j^{'}}\left(C, \phi\left(\gamma_C^{'}\right)\right)$. □

Proof of Theorem 3.

Proof. By Lemma 1 $U_x(N, \gamma^*) = U_x\left(C\phi\left(\gamma_C^{'}\right)\right)$. Notice that $\gamma_{x\in C}^{'} < \gamma_{x\in N}^{*}$, otherwise $\frac{\gamma_x^*}{\sum \gamma_{i\in N}^*} = \frac{\gamma_x^{'}}{\sum \gamma_{i\in C}^{'}}$, $C = N_+$thus is not a proper sub-coalition. Consequently, $\gamma_C^{'}$ is adjustable after the small shocks $0 < \eta < \underline{\eta_0} = \gamma_{x\in N}^{*} - \gamma_{x\in C}^{'}$.

There are two cases to consider.

First, if $\left(C, \gamma_C^{'}\right)$ is the only deviation coalition whose deviation constraint binds. That is, for any winning sub-coalition $\widetilde{C} \neq C$ such that $\sum \gamma_{i\in\widetilde{C}}^{'} > \sum \gamma_{i\in N\backslash\widetilde{C}}^{*}$, $\gamma_{\widetilde{C}}^{'} \leq \gamma_{\widetilde{C}}^{*}$, we have $U_j(N, \gamma^*) > U_j\left(\widetilde{C}, \phi\left(\gamma_{\widetilde{C}}^{'}\right)\right)$ for some $j \in \widetilde{C}$. Then for small enough grid ν, after a small negative shock hits x, we still have we have $U_j(N, \gamma^{*,\eta}) > U_j\left(\widetilde{C}, \phi\left(\gamma_{\widetilde{C}}^{'}\right)\right)$, where $\gamma^{*,\eta} = \left(\gamma_x^* - \eta, \gamma_{-x}^*\right)$. This means that $\left(C, \gamma_C^{',\eta}\right)$ is the only stable improving sub-coalition. Thus agents in $N\backslash C$ has to burn additional power to accommodate x, until $U_x(N, \gamma^*) = U_x\left(C, \phi\left(\gamma_C^{'}\right)\right)$. In particular, denote the set of $\widetilde{C}$ as $\widetilde{\mathcal{C}}$, denote $\eta_{\widetilde{C}}$ as the largest shock that makes $U_{j\in N}(\gamma^{*,\eta}) \geq U_{j\in\widetilde{C}}\left(\phi\left(\gamma_{\widetilde{C}}^{'}\right)\right)$ for some $j \in \widetilde{C}$, we can calculate $\underline{\eta_1} = \min\left\{\eta_{\widetilde{C}}\right\}_{\widetilde{C}\in\widetilde{\mathcal{C}}}$.

Second, if there are other deviation coalitions whose deviation constraint binds. There are two scenarios: 1. the binding agent is not x. Then after the shock the deviation constraint is slack. We go back to the first case; 2. the binding agent is x, then we calculate the upper bound of the shock size by $\underline{\eta_2} = \min\left\{\gamma^*_{x\in N} - \gamma'_{x\in\widehat{C}}\right\}_{\widehat{C}\in\widehat{\mathcal{C}}}$, where $\widehat{\mathcal{C}}$ is the set of deviation coalitions with binding agent x.

Lastly we denote $\underline{\eta} = \min\left\{\underline{\eta_0}, \underline{\eta_1}, \underline{\eta_2}\right\}$. $\square$

Proof of Theorem 4.

Proof. We once again try to show that $\left(C, \gamma_C^{',\eta}\right)$ remains the unique stable improving sub-coalition after the shock. The proof is similar with that of Theorem 3. A small positive shock from the agents in A makes $\left(C, \gamma'_C\right)$ feasible but keeping incentive constraints of other winning coalitions slack. Denote the threshold derived from such process as $\overline{\eta_1}$. The next step is to check whether an empowered agent j brings additional stable improving coalitions to consider. Pick $\overline{\eta_2} < \gamma_j - \gamma_j^*$, suppose there is such a new stable improving coalition $\widehat{C}$, where $j \in \widehat{C}$, $\widehat{\gamma}_j < \gamma_j$, then we know that for agent j, γ_j^* is a better response than $\widehat{\gamma}_j$. Therefore $\left(C, \gamma_C^{',\eta}\right)$ remains the unique stable improving coalition to consider. Lastly we denote $\overline{\eta} = \min\left\{\overline{\eta_1}, \overline{\eta_2}\right\}$. $\square$

Appendix B. Additional Examples

Suppose that in the burning meeting an agent can only adjust his own power downwards, i.e. $B_i \equiv \{i\}$, the following example shows that there exists a equilibrium which includes a stable improving sub-coalition.

Consider a committee with 4 members. Examples can be the Board of Directors in a corporation or standing members in national security council. The committee decides a policy which generates rent of 1 to share among the members. The initial formal power structure is $\gamma = (\gamma_x = 4, \gamma_y = 4, \gamma_z = 4, \gamma_w = 6)$, where we refer to x, y, z, w as four committee members, with w being the leader of the committee, or the "first among the equals", an expression used from Joseph Stalin to Saddam Hussein. The agenda is top-down, that is, w moves first, followed by z, y and x. As in Example 1, members try to form stable improving alliances to increase their relative power. A feasible option can be among y, z and w, with underlying new power balance $\gamma_{C'} = \left(\gamma_y^{'} = 3, \gamma_z^{'} = 3, \gamma_w^{'} = 6\right)$, which itself is a equilibrium . However, to achieve this, both y and z has to adjust power. Suppose y adjusts first in the transition to $\gamma_z^{'} = 3$. When it is z's turn to adjust, he faces power structure $\gamma^{'} = (4, 4, 3, 6)$, which he now has an opportunity to seek for a new alliance with w *only*, which generates $\gamma_{C"} = \left(\gamma_y^{"} = 4, \gamma_w^{"} = 4\right)$ and an eventual equilibrium of $(0, 4, 2, 6)$. Agent y enjoys more rents in inner circle $C^{'}$ than in $C"$ since $\frac{4}{12} > \frac{3}{12}$. Therefore the transition cannot continue. Furthermore, anticipating y's incentive to switch alliances (and leaves z with rents $\frac{2}{12}$), z no longer wishes to initiate power struggles that centers around $C^{'}$. Thus the original power structure stays stable.

This is a key difference of the sequential coalition formation game comparing with standard cooperative coalition formation. In cooperative game literature, it is assumed that upon a rent-improving allocation the involving agents can commit to form a coalition to collect the rents. It is, however, not the case in this model, because during the formation of such coalition, some agents might have received higher rents in the transition when other agents burn power, and consequently choose to divert to another deviation which may be not desirable from the perspective of the initiator. Such motive provides an additional economic force to support a wider range of institutional arrangement. In the literature, this is closest to Chwe (1994) in the discussion of the credibility of selecting a deviation strategy.