

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Journal of Comparative Economics

journal homepage: [www.elsevier.com/locate/jce](http://www.elsevier.com/locate/jce)

## The political agenda effect and state centralization

Daron Acemoglu<sup>a</sup>, James A. Robinson<sup>b</sup>, Ragnar Torvik<sup>\*c</sup><sup>a</sup> *Massachusetts Institute of Technology, Department of Economics, E52-380, 50 Memorial Drive, Cambridge MA 02142, USA*<sup>b</sup> *Harris School of Public Policy, University of Chicago, 1155 East 60th Street, Chicago, 60637, IL USA*<sup>c</sup> *Department of Economics, Norwegian University of Science and Technology, Trondheim N-7491, Norway*

## ARTICLE INFO

## Keywords:

Conflict  
Escalation effect  
Political agenda effect  
Public good provision  
State capacity  
State centralization

## JEL classification:

D70  
H11  
P48

## ABSTRACT

We provide a potential explanation, based on the “political agenda effect”, for the absence of, and unwillingness to create, centralized power in the hands of a national state. State centralization induces citizens of different backgrounds, interests, regions or ethnicities to coordinate their demands in the direction of more general-interest public goods, and away from parochial transfers. This political agenda effect raises the effectiveness of citizen demands and induces them to increase their investments in conflict capacity. In the absence of state centralization, citizens do not necessarily band together because of another force, the escalation effect, which refers to the fact that elites from different regions will join forces in response to the citizens doing so. Such escalation might hurt the citizen groups that have already solved their collective action problem (though it will benefit others). Anticipating the interplay of the political agenda and escalation effects, under some parameter configurations, political elites strategically opt for a non-centralized state. We show how the model generates non-monotonic comparative statics in response to the increase in the value or effectiveness of public goods (so that centralized states and public good provision may be absent precisely when they are more beneficial for society). We also suggest how the formation of a social democratic party may sometimes induce state centralization (by removing the commitment value of a non-centralized state), and how elites may sometimes prefer partial state centralization.

## 1. Introduction

There is a great deal of variation across societies both historically and today in the degree to which a national state has achieved the Weberian monopoly of violence over its territory, developed the authority and the capacity to enforce laws, maintain law and order, and raise taxes and provide public goods — a vector of attributes that social scientists call “state capacity”. A growing literature has documented the importance of state capacity for economic outcomes (e.g., Johnson, 1982; Amsden, 1989; Evans, 1995; Evans and Rauch, 1999, 2000; Besley and Persson, 2009, 2011; Acemoglu et al., 2015, 2016, and Acemoglu and Robinson, 2019).

Fundamental for establishing capacity is whether or not a state has managed to centralize authority and moved from various systems of “indirect rule” to a situation where a national state actually directly organizes these activities. The extent to which this process, which we refer to as *state centralization*, has been undertaken, varies greatly as well. At one end of the scale there are countries such as most Western European ones, as well as China and Japan, with a high degree of state centralization, while at the other end, the Afghan, Somali, Pakistani, Philippine, and Colombian states, among many others, are very far from having forged such

\* Corresponding author.

E-mail addresses: [daron@mit.edu](mailto:daron@mit.edu) (D. Acemoglu), [jamesrobinson@uchicago.edu](mailto:jamesrobinson@uchicago.edu) (J.A. Robinson), [ragnar.torvik@ntnu.no](mailto:ragnar.torvik@ntnu.no) (R. Torvik).<https://doi.org/10.1016/j.jce.2020.03.004>

Received 4 February 2020; Received in revised form 25 March 2020; Accepted 27 March 2020

Available online 21 May 2020

0147-5967/ © 2020 The Authors. Published by Elsevier Inc. on behalf of Association for Comparative Economic Studies.

centralization. In non-centralized states, rule and authority are delegated to other entities, such as traditional elites in the Philippines, tribal elites in Pakistan and various types of warlords in Colombia. Though state centralization appears to be a critical prerequisite for establishing capacity,<sup>1</sup> and is mostly taken for granted (e.g. in the literature on East Asian development), we are far from a consensus as to why many states have not centralized power or even attempted to establish the monopoly of violence over their territories. This question is particularly puzzling since it would appear that all power-holders should want to monopolize power in their countries (e.g., North, 1981, Chapter 3). If so, why is the state so hard to centralize?

This paper investigates political economic causes of lack of state centralization — and thus one of the factors keeping state capacity perennially low in many societies. We model state centralization as the decision of different elites to coordinate their actions and delegate their control over (local) violence to a national state. Critical to our explanation is the *political agenda effect*, which is the name we give to the possibility that state centralization changes the dynamics of political action and conflict in society. The anticipation of the political agenda effect may discourage elites from coordinating their actions and centralizing the state. More specifically, when citizens from different regions, sectors, interests, backgrounds, or ethnicities organize jointly, their agenda will change in a direction that makes their demands from power-holders (elites) focus more on (general-interest) public goods. The greater efficiency of public goods — relative to transfers — encourages them to invest more in their conflict capacity, increasing the effectiveness of their demands.<sup>2</sup> In turn, state centralization induces citizens to organize nationally as well — rather than at the local or the ‘parochial’ level. It is this indirect effect of state centralization that makes elites often prefer a non-centralized state. Herein lies the main mechanism of our model: the elites may strategically opt for a non-centralized state so as not to induce the citizens to organize nationally and thus avert the political agenda effect.

The process of centralization and strengthening of the national state in Britain between 1758 and 1834 (see, e.g., Brewer, 1990; Harding, 1996) illustrates the political agenda effect — the changes in the societal equilibrium accompanying state centralization. Charles Tilly’s (1995) classic study emphasizes that this process “... brought eighteenth century Britons into open confrontation with one another...” (p. 5). At the start of the period, Tilly notes, contention was about

“local people and local issues, rather than nationally organized programs and parties” (p. 5),  
[but] “between 1758 and 1833 a new variety of claim-making had taken shape in Britain... Mass popular politics had taken hold on a national scale” (p. 13).

Tilly observes how the forms of collective action that emerged were completely new. For example, the open meeting became “a kind of demonstration... a coordinated way of publicizing support for a particular claim on holders of power. Frequently a special purpose association, society or club called the meeting. What is more, meetings recurrently concerned national issues, emphatically including issues that the government and Parliament were on their way to deciding” (p. 10). Tilly further points out that “the means by which ordinary people made collective claims... underwent a deep transformation: increasingly they involved large scale, coordinated interaction that established direct contact between ordinary people and agents of the national state” (p. 14).

Tilly also argues that the driving force of this changing nature of contention and increasing coordination of civil society was indeed the development of the national state:

“an expansion of taxes, national debt, and service bureaucracies, which increased not only the state’s size but also its weight within the economy.. These changes... promoted a shift towards collective action that was large in scale and national in scope” [and] “the expansion of the state pushed popular struggles from local arenas and from significant reliance on patronage towards autonomous claim-making in national arenas” (Tilly, 1995, p. 49, 53).

This was precisely because, according to Tilly, the state gained “increasing importance... for the fates of ordinary people” (p. 16), and that this<sup>3</sup>

“generated threats and opportunities. Those threats and opportunities in turn stimulated interested parties to attempt new sorts of defense and offense; to match association with association, to gain electoral power, to make direct claims on their national government. Through long strenuous interaction with authorities, enemies and allies, those ordinary people fashioned new ways of acting together in their interests and forced their interlocutors to change their own ways of making and responding to claims. Cumulatively, struggles of ordinary people with power-holders wrought great changes in the British structure of power” (Tilly, 1995, p. 16).

Our model formalizes and further elucidates these interactions by considering a society with  $N$  regions, with location being the only dimension of heterogeneity across groups of citizens.<sup>4</sup> Each region is also inhabited by a group of elites. Citizens can make demands from elites backed up by their ability to engage in conflict. If they are not able to engage in conflict, the elites will not

<sup>1</sup> For the link between different aspects of state centralization and the capacity of the state to effectively provide public goods and regulate economic activity, see the historical accounts by Jean-Laurent (1990), Epstein (2000) and Nye (2007), and empirical work by Dincecco and Katz (2016), Gennaioli and Rainer (2007), Michalopoulos and Papaioannou (2013), and Osafo-Kwaako and Robinson (2013).

<sup>2</sup> Here ‘citizens’ stands for members of civil society, distinguished from those who are the political elite or the power-holders (or their direct agents, such as the police or the military). The citizens could be acting as peasants, workers or civil society members in formulating demands and participating in potential conflict with the state and its agents.

<sup>3</sup> Johnson (2015) presents a similar argument to Tilly’s in the French case.

<sup>4</sup> This is for specificity, and working with other dimensions of heterogeneity would lead to essentially identical results.

respond to their demands and provide any redistribution. When they are able to engage in conflict, the elites will placate their demands with the cheapest form of redistribution — consisting of a combination of direct transfers and public good provision — to convince them not to engage in costly conflict.

The effectiveness of the demands of citizens is determined by two factors: they need to invest in their conflict capacity, which is costly, and moreover, only some groups of citizens (in our model, for simplicity, citizens from one region only) are able to solve their collective action problem and invest. These “strong” citizens then have a choice — either engage in local demands (backed up by the threat of local conflict), or organize other “weak citizens” and engage in demands and conflict at a national (or at the very least supra-local) level. In our model, when the demands are local, the cheapest way of placating them for the local elite is to make parochial, local transfers. However, when the demands are at the supra-local (or national) level, general-interest public goods become a better option, because different types of public goods benefit all citizen groups — not just the local group. This formalizes the political agenda effect: when the conflict is at the supra-local level, citizens will invest more in their conflict capacity because they recognize that a successful outcome in the conflict will lead to public good provision, which is more beneficial for them (and in the absence of redistribution from the elite, they themselves will be able to invest in public goods in case they win the conflict). This argument further shows that the political agenda effect also provides a mechanism for why, as documented by the studies cited in footnote 1, public good provision will be associated with centralized states — parochial transfers emerge as the most economical way of meeting citizen demands in the absence of state centralization.

Weighing against a national organization, however, is the *escalation effect*: if the citizens band together in a national organization, this will escalate the fight by inducing the elites to also form a national organization and pool their resources to fight against the citizens. The escalation effect features prominently in the calculus of strong citizens: by forming a national organization, they will directly benefit the weak citizen groups (who would have otherwise remained unorganized), but they may face a lower probability of success and thus lower transfers because of the escalation of the conflict.<sup>5</sup>

We show that for an interesting part of the parameter space, in the absence of a centralized state, the escalation effect is potent enough that strong citizens do not initiate a national organization, and as a result, the elites are able to avoid the political agenda effect. Crucially, if the elites were to choose a centralized state, this would induce citizens to also join up in a national coalition, putting in motion the political agenda effect. As a result, the elites strategically choose a non-centralized state in order to avoid such political escalation, and prefer to deal with demands from their local citizens, if any, by making parochial transfers.<sup>6</sup>

Several important, and in some cases surprising, conclusions follow from this theory. First, in the relevant region, citizens benefit *ex ante* from a centralized state, because it enables all of them to organize and make demands, and as already noted, these demands will be met via the provision of general-interest public goods. In contrast, with a non-centralized state, only strong citizens are able to do so (and they do not internalize the positive impact they have on other citizen groups). Second, and paradoxically, a greater value of public goods — and similarly, lower heterogeneity in the preferences of citizens — can make the provision of public goods and state centralization less likely. This is because more valuable public goods, by making citizens more likely to coordinate under state centralization, may further discourage elites from building a centralized state, thus push in the direction of parochial, location- or issue-specific transfers. Third, we show how a social democratic party may change the nature of equilibrium. This happens when such a party induces citizens to band together before the identity of the strong group is revealed, wresting away from the elites the first mover advantage (which enabled the elites to effectively commit to not banding together as long as the citizens did not do so also). Put differently, once citizens commit to acting in a nationally coordinated manner, the strong citizen groups will always organize the weak groups, and in response, elites now prefer the centralized state. In the relevant part of the parameter space, this option always increases the *ex ante* utility of citizens. Finally, we show how elites may opt for partial state centralization, which enables them to increase their power in the conflict and thus reduce the transfer they need to make, while still making use of the escalation effect to discourage strong citizens from forming coalitions with weak citizens.

The escalation and political agenda effect are evident in many experiences of political centralization, particularly in Post-World War II Southeast Asia. For example, Malaysia was split by the British prior to World War II into the Crown colony of the Straights Settlement, (consisting of Singapore, Melaka, Penang and Province Wellesley), the Federated Malay States and the Unfederated Malay States. The states were ruled indirectly through the traditional rulers (Emerson, 1937). In addition there was a Chinese Protectorate which dealt with any issues related to the Chinese people. This patchwork of polities meant that Malaya was politically highly non-centralized (e.g., (Andaya and Andaya, 1982, p. 245)). This changed after World War II. During the war, the Japanese took over Singapore as a colony and united the rest of the country as a protectorate, weakening traditional rulers and fomenting a Malay national identity in opposition to the British. Andaya and Andaya argue that “Malays increasingly began to see themselves as belonging to a Malaya-wide entity, rather than to their individual states,” (1982, p. 248), which was a very different situation from the types of parochialism evident previously. This national identity came together with the rapid growth of the Malayan Communist Party (MCP), which was armed during the war by the British to fight the Japanese, and thereafter assumed virtual *de facto* control of the country when the Japanese surrendered. The response of the British was to propose to merge all of the polities into a Malayan Union in a way that implied equal treatment for Malays and Chinese. But this policy, in turn, triggered a response akin to our political agenda effect: “For the first time in history, the Malays rose in one movement to fight against the formation, putting aside parochial

<sup>5</sup> Such escalation may not always harm the strong citizens, but will always do so at the *margin* when they are indifferent between engaging in conflict by themselves and forming a broader coalition.

<sup>6</sup> This strategic lack of state centralization may further contribute to the low capacity of the state: the elites do not provide public goods, and do not need to build the capacity of the state to increase the efficiency of public goods.

sentiments relating to individual states, districts or clans” (Hooper, 2003, p. 8). This reaction led to a compromise Federation of Malaya, and to strengthened state institutions to contain the Communist rebellion. (Harper, 1999, pp. 195-196), notes that “During the Emergency the classic functions of the state — military, fiscal, administrative — were greatly extended and new ones adopted. A centralized federal government grew in strength — The state became for the first time a physical presence in the lives of many of its subjects.”. Slater’s summary of the situation thus emphasizes how the elites had to centralize the state in response to this bottom-up conflict. He states: “Malaysia’s strong central state” has its roots in “elite responses to especially challenging pressures from below” (Slater, 2010, p. 59), and

“endemic and unmanageable threats from below inspired the construction of a strong and centralized state apparatus in Malaya in the decade following World War II. By the time of independence in 1957, the Malayan state was already noteworthy for the effectiveness of its coercive and administrative institutions. The initial processes of state-building were compounded and accelerated in the early 1970s, as the racial riots of May 1969 provided a powerful impetus for government leaders to strengthen their coercive grip and increase their fiscal demands upon the Malaysian population” (Slater, 2010, p. 147).

The situation in Indonesia in the 1960s was also very similar. Once again, the significant strengthening of state institutions (in the context of the transition from Sukarno to Suharto and the emergence of the so-called New Order, e.g., Anderson 2001) came in response to the communist insurgency. Slater’s assessment of this case further illustrates our escalation effect: “it was the dramatic rise of contentious class politics in the mid-1960s, via the mobilization of a powerful, grassroots communist party with a massive rural and urban membership, that spurred a remarkable increase in elite collective action upon the birth of the Suharto regime” (2010, pp. 26-27). He then generalizes these two cases to the entirety of Southeast Asia, again underscoring our escalation effect: “Mass movements preceded the rise of authoritarian Leviathans... New elite coalitions arose in active support of... increased state centralization” (Slater, 2010, p. 23).

That the anticipation of the political agenda effect can inhibit state-building is well illustrated by the history of European colonial rule as well. Most colonial powers, like the British ruled “indirectly” and delegated a large number of tasks, such as taxation and the provision of justice, to local traditional rulers. They saw little benefit from providing public services, and all the dangers of building a capable administration, which would just increase demands from society and put further roadblocks on their plans to run a profitable colony. Sir Donald Cameron, the governor of Tanganyika in the 1920s, clearly identified the political agenda effect when he noted

“If we set up a European form of administration, the day will come when the people of the Territory will demand the British form of administration shall pass into their hands... If we aim at indirect administration through the appropriate Native Authority... we shall be building an edifice... capable of standing the shock which will inevitably come when the educated native seeks to gain possession of the machinery of government and to run it on Western lines... If we treat them properly, moreover, we shall have the members of the Native Administration on our side” (Iliffe, 1979, p. 322).

Hechter (2000) notes more broadly how shifts from indirect to direct rule led to large reactions from society in the face of much more direct centralized administration, with this reaction often taking the form of nationalism.

Our paper is related to the growing literature in economics and political science on the role of state capacity, political centralization and the formation of the state, mentioned already above. Some of this literature has developed political mechanisms that deter elites from building states. Acemoglu (2005) suggests that states with strength beyond a certain level will make citizens worse off and may be resisted (unless a “consensually strong state” that is effectively monitored by the citizens emerges). Besley and Persson (2009), Besley and Persson (2011) emphasize that if incumbent elites are threatened with the loss of power then they may refrain from building a state because the capacity can be subsequently used against them. Our model develops a very different mechanism, with different predictions. For instance, in these previous studies, when public goods become more valuable, this makes it more attractive to build a state, but this is not necessarily the case in our model. Acemoglu et al. (2011) develop a model where incumbent elites face democratization and create an inefficient state in order to favorably influence the democratic political equilibrium. The large literature on civil war can also be interpreted in terms of state formation, for example political factors may deter states from eliminating rebel groups and establishing a monopoly of violence (e.g., Acemoglu et al., 2010; Acemoglu et al., 2013).

Our results on the political agenda effect are also related to the large literature on clientelism which has emphasized how politicians target transfers to their supporters (Bates, 1981; Shefter, 1977, 1993; Lizzeri and Persico, 2001; Kitschelt and Wilkinson, 2007; Robinson and Verdier, 2013; Stokes et al., 2013) and to the long-standing puzzle in political science of when politics focus on the provision of general-interest public goods as opposed to patronage, clientelism and parochial benefits (see Kitschelt, 2000, for an overview). We provide a new argument here based on the political agenda effect — public goods politics emerge when citizens organize collectively, a process which leads to a demand for public as opposed to parochial transfers. This argument also provides a potential explanation for the findings of Anderson et al. (2015), which document how local elites in Maharashtra, India are able to dominate politics and curtail the provision of public goods, among other things, by clientelism, particularly aimed at preventing coordination by non-elite citizens. Our main result is related to, but distinct from, Lizzeri and Persico’s (2004) argument that when politicians need to appeal to a larger number of voters (due to democratization), it becomes more cost-effective for them to do this by providing public goods. Our emphasis on the roles of the political agenda effect and state centralization in curtailing clientelism is also different from one of Shefter’s (1977) suggestions that clientelism is weakened when new political parties mobilize outside the

existing political system.<sup>7</sup>

Several strands of the literature on state formation in the sociology and political science literatures are connected to our theory as well. One branch emphasizes the role of social movements, which the state may influence by using its resources or by other means (Tilly 1978), or by favoring some specific groups, for example, through selective policy or repression (McAdam et al. 1988). Another important branch, also due to Tilly (1990), emphasizes the connection between war-making and state-building. According to this argument, bureaucratic state structures emerge in order to coordinate taxation, conscription and requisitioning necessary for large armies. Wittfogel's (1957) famous Oriental despotism thesis can be viewed as a precursor of this argument. Wittfogel claimed that early state formation originated from the desire to establish large-scale irrigation systems in places where non-irrigated agriculture was infeasible or less productive. Yet another branch builds on Scott's (1998) seminal work highlighting efforts of states to make society "legible", for example, by collecting census data, imposing uniform language and standardizing weights and measures. Several other scholars also emphasize that state formation or centralization is specifically motivated by the desire to control society, such as in Anderson's (1974) and Hechter and Brustein's (1980) theories of the emergence of absolutism in early modern Europe, or Saylor's (2014) examination of contemporary state-building in several developing countries. Our approach is complementary to all of these arguments, but emphasizes how state efforts to gain control over society, regardless of their exact motivation, tend to trigger the political agenda effect and thus may sometimes end up threatening.

Finally, our model is connected to the literature on federalism as well. The seminal work by Riker (1964) argued that federalism emerged from a bottom-up process of autonomous political units ceding authority to a federal government to exploit the potential benefits of coordinating on such things as national defense. Subsequent research has focused on how federal systems can deal with specific challenges, for example externalities and public good provision (e.g., Lockwood, 2002; Besley and Coate, 2003), and with the exception of Crémer and Palfrey (1999), there has been little work on the endogenous emergence of federalism. Closer to our paper are models of the circumstances under which federal systems may decentralize authority or even break into component parts (e.g., Bolton and Roland 1997; Filippov et al. 2004; de Figueiredo and Weingast 2005; O'Neill 2005). Though such positive theories can explain why centralization may not occur, we are not aware of other works that have studied the impact of centralization on the organization of society or argued that this is an important disincentive to centralization.

The rest of the paper is organized as follows. In Section 2 we introduce our model of state centralization. In Section 3 we study the equilibrium political power and political agendas with and without state centralization. The equilibrium emergence of state centralization is discussed in Section 4. In Section 5 we extend the model to deal with a case where the citizens can coordinate into a national political movement ahead of an eventual centralization of the state, and in Section 6 we extend the model to discuss a case where the elites may centralize the state only in parts of the territory. Section 7 concludes. We present some omitted proofs and discuss several additional extensions of the model in the Appendix.

## 2. Model

In this section, we present our basic model of state centralization. The ideas discussed in the Introduction are conceptualized in the context of a model consisting of regional heterogeneity though, as noted there, other dimensions of heterogeneity would be entirely analogous.

### 2.1. Preferences and technology

We consider a society consisting of  $N$  regions, and we use  $i \in N \equiv \{1, \dots, N\}$  to denote a particular region. Each region is inhabited by a set of homogeneous citizens and homogeneous elites, each with measure normalized to 1. Throughout, there will be no conflict of interest among citizens or elites within a given region, and we will not distinguish between the group and a particular element thereof, and use the superscripts  $c$  (respectively,  $e$ ) to denote the entire group of citizens (respectively, elites) or an individual member.<sup>8</sup>

Elite preferences depend only on their consumption, denoted for an elite agent of region  $i$  by

$$C_i^e,$$

where  $C_i^e \geq 0$ . Since the total measure of elites within each region is normalized to 1,  $C_i^e$  also denotes total consumption expenditure of elites in region  $i$ .

Citizen preferences depend both on their consumption and on public goods (such as public schooling or the quality of roads, which may matter less for elites who are able to afford their own private alternatives). However, reflecting the potential conflict of interest across regions, the quality of these public goods within an individual's own region matters more for her than those in other regions. Namely, the utility of a citizen from region  $i$  is

<sup>7</sup> The emphasis on the state shapes society and vice versa is also related to the work of Habermas (1991), who suggested the notion of a 'public sphere' as an inclusive place in society where people come together to discuss and deliberate and form opinions. Habermas viewed this as related to state formation, noting that "Civil society came into existence as the corollary of a depersonalized state authority" (1989, p. 19). Other scholars, such as Katznelson (1985), Evans (1995), and Migdal (1988, 2001), have also emphasized the interaction between the state and society, but have tended to treat both the strength of the state and society as historically determined.

<sup>8</sup> As noted above, 'region' here stands for either locational heterogeneity, ethnic or religious heterogeneity, or heterogeneity in terms of other preferences. An important application of the model is to ethnic heterogeneity, which would imply that the conflict in the non-centralized state is between elites and citizens of a certain ethnicity, and state centralization involves elites of different ethnicities banding together.



$$C_i^c + G_i + (1 - \lambda) \sum_{j \neq i} G_j,$$

where  $C_i^c \geq 0$  is the private consumption of citizen  $i$ ,  $G_i \geq 0$  is the total quantity of public goods in the individual’s own region, and  $\lambda \in [0, 1]$  parameterizes the extent of heterogeneity in preferences among citizens: when  $\lambda$  is close to zero, an individual cares about public goods in other regions equally (e.g., because this facilitates trading or an individual can get easy access to these public goods), and conversely when  $\lambda$  is close to 1, an individual only cares about public goods in her region.

Total output of the consumption good within each region is  $Y$ , and we simplify the setup by assuming that this is inelastically produced and initially accrues to the elite (e.g., it is their endowment of land or natural resources). It can also be taxed without any distortions. One unit of this consumption good can be converted into  $\mu$  units of any of the regional public goods.

Let us next define, for future reference,  $\Phi(n)$  as the *marginal utility of funds* for citizens. Namely, this is the maximum symmetric citizen utility that a coalition of citizens from a coalition  $N'$  with  $|N'| = n$  can achieve from one unit of the consumption good per region. To compute this, note that if we convert a fraction  $x$  of the unit of consumption from each region into the public good from that region (by the symmetry requirement), then each citizen will have a utility of  $(1 + (1 - \lambda)(n - 1))\mu x + 1 - x = ((1 - \lambda)n + \lambda)\mu x + 1 - x$ . Clearly, this expression always has a corner maximizer in  $x$ , thus enabling us to write

$$\Phi(n) \equiv \max\{1, ((1 - \lambda)n + \lambda)\mu\}. \tag{1}$$

It is also straightforward to see that  $\Phi(n)$  is nondecreasing in  $n$ . We next impose our first parametric assumption on this  $\Phi$  function:

**Assumption 1.**  $\Phi(1) = 1$  and  $\Phi(N) > 1$ .

The first part of this assumption imposes that  $\mu < 1$ , which ensures that when in isolation, a single group prefers not to invest in the public good. The second part implies that when all  $N$  regions are combined, it is worthwhile to invest in public goods. In particular, the second part requires that  $\lambda$  is not too large. Substantively, this assumption restricts attention to situations in which the demand for public goods will be greater when all regions are simultaneously investing in public goods.<sup>9</sup> This assumption thus restricts attention to the part of the parameter space that is of interest for our analysis. Since  $\Phi(n)$  is nondecreasing, Assumption 1 also implies that there exists a unique  $n^*$  such that  $\Phi(n) > 1$  for  $n > n^*$  but  $\Phi(n^*) = 1$ .

**Remark 1.** Our analysis below will show that the functional form of  $\Phi(n)$  plays no major role in our results. Thus we could generalize (1) to

$$\Phi(n) = f(n)\max\{1, ((1 - \lambda)n + \lambda)\mu\},$$

where  $f(n)$  is a nondecreasing function reflecting the greater effectiveness of using funds when resources are deployed at the more centralized level or in a more coordinated fashion. For instance, the case where  $f(n) = 1$  for all  $n < N$ , and  $f(N) > 1$  can be interpreted as capturing the greater efficiency of a “centralized state” allocating funds for all regions.

## 2.2. Policies, political power and state centralization

Policies in this economy concern how much of each region’s output  $Y$  will be taxed and how much of this will be provided as direct transfers to citizens and how much of it will be invested in public goods. These policies are decided by the group which has local or national political power. Initially, political power in region  $i$  rests with the elite from that region, but may be contested by citizens. We next describe how this conflict takes place and the technology for conflict. The key is whether the state is “centralized”. As described in the Introduction, our focus is whether political power and fiscal policy are determined entirely at the local level or are centralized to the national level.

The two cases we initially consider are total lack of state centralization, denoted by  $s = 0$ , and full centralization, analogously denoted by  $s = 1$ . Under a non-centralized state, each local elite acts entirely autonomously, without any coordination, whereas under full centralization, they commit to transfer power to a national political body that represents their collective preferences as we describe next.

Under both centralized and non-centralized state structures, citizens can contest political power. We assume that the extent to which they can do so depends on whether they are able to solve their within-region collective action problem, and for simplicity we assume that only one of the  $N$  regions (drawn uniformly at random) will be able to do so, and the remaining  $N - 1$  regions will not. We refer to the citizens that have solved their collective action problem as “strong”, and the citizens in the remaining regions as “weak”. Strong citizens can contest local political power, while weak citizens cannot unless they join up in a coalition or “organization” with the strong region.<sup>10</sup>

Suppose, without loss of any generality, that it is citizens in region 1 that are strong. An additional decision for this group of citizens is whether to form a coalition with other regions. We denote by  $Z^c = 0$  the decision not to form such a coalition, and by  $Z^c = 1$  the decision to offer to form a coalition to citizens from other regions. We ignore for now the decision to offer to form a

<sup>9</sup> Many public goods, such as infrastructure or public health investments, would have this property.

<sup>10</sup> The interpretation here is that when weak groups are part of an organization with a strong group that has already solved its collective action problem, they can also de facto solve their own collective action problem, thus becoming capable of contesting power in their region or in a national conflict.

coalition with a subset of this homogeneous set of citizens from other regions; we return to this issue in the Appendix and show that this simplification is without consequence. Following the choice of  $Z^c = 1$ , citizens from all other regions decide whether to join this coalition, denoted by  $z_i^c \in \{0, 1\}$  for  $i \neq 1$ . Let us also designate  $N^{z^c=1} \equiv \{i: i = 1 \text{ or } z_i^c=1\}$  (and note that region 1 is always in  $N^{z^c=1}$ ). If  $Z^c = 0$ , it means that citizens from region 1 will engage only in local conflict and present local demands from their elites. If  $Z^c = 1$  and  $z_i^c = 1$  for all  $i \geq 2$ , then citizens from different regions will have formed a national organization, and engage in national conflict and present national demands.

Under a non-centralized state ( $s = 0$ ), after observing the realization of the strength of citizens of different regions and  $N^{z^c=1}$ , each regional elite also decides whether to join up in a coalition. We use a similar notation,  $z_i^e$ , to denote the decision of the elite from region  $i$  to form a coalition with the elite from region 1 (ignoring coalitions excluding region 1 is without loss of any generality as will become apparent). We denote by  $N^{z^e=1} \equiv \{i: i = 1 \text{ or } z_i^e=1\}$  the coalition of the elite.

Under a centralized state ( $s = 1$ ), on the other hand, the elites pool their resources and delegate these to a national organization, which then confronts all demands and conflict from the citizens.

Subsequent to the state centralization decision and the coalition formation decisions, each region within this citizen-side coalition decides how much to invest in the conflict technology (e.g., armaments or organizational capital), denoted by  $\theta_i^c$ , with the collection of these investments being  $\{\theta_i^c\}_{i \in N^{z^c=1}}$ . We assume that the cost of investment in terms of the final good is given by  $\Gamma(\theta_i^c)$ , which is continuously differentiable and satisfies  $\Gamma(0) = 0$ ,  $\Gamma'(\theta_i^c) > 0$ ,  $\Gamma''(\theta_i^c) \geq 0$  for all  $\theta_i^c \geq 0$ , and  $\lim_{\theta_i^c \rightarrow \infty} \Gamma'(\theta_i^c) = \infty$ . Each regional elite has conflict capacity given by  $\theta^e \geq 0$ . We take this elite-side capacity as exogenous to simplify the discussion and show in the Appendix that endogenizing it does not affect our main results.

Finally, each regional elite facing the threat of conflict — i.e., those with indices belonging to the set  $N^{z^c=1} \cup N^{z^e=1}$  — decides on a transfer-public good package to encourage peaceful settlement with the citizens. We denote the package offered by elites in region  $i$  by  $(T_i, G_i)$ , where  $T_i + G_i/\mu \leq Y$  (with  $Y - T_i - G_i/\mu$  being left for the consumption of the elite in region  $i$ ). No concession can be simply captured by setting  $(T_i, G_i) = (0, 0)$ . More specifically, there are three possibilities to consider. Either (i)  $N^{z^e=1} = \{1\}$ , in which case the elite in region 1 individually offer a policy package. Or (ii)  $N^{z^e=1} \supset \{1\}$ , in which case the elites in this coalition jointly decide on a policy package, which they will each offer.<sup>11</sup> The fact that they all offer the same policy package is natural, since at this point, there is no conflict of interest among regional elites in  $N^{z^e=1}$  for in case this offer is rejected, they will all have exactly the same probability of losing the conflict and suffering the same consequences (as we describe next). Or finally (iii)  $i \in N^{z^c=1} \setminus N^{z^e=1}$ , in which case this elite group is facing the organized coalition  $N^{z^c=1}$  but is not part of  $N^{z^e=1}$ , so will have to make an individual offer again. Following the offer  $(T_i, G_i)_{i \in N^{z^c=1}}$  or  $(T_i, G_i)_i$ , the corresponding group of citizens decide whether to accept this concession or to fight ( $f^e_{N^{z^c=1}} \in \{0, 1\}$  or  $f_i^e \in \{0, 1\}$ ). If  $f^e = 1$ , there will be fighting, and the winner of the conflict is determined stochastically.

To explain how conflict takes place, first consider the case in which  $s = 0$  (i.e., no state centralization) and  $N^{z^e=1} = N^{z^c=1} = \{1\}$ , so that conflict is local. Then there will be a single conflict in region 1, and no conflict with any other region, since all other citizen groups are weak. The outcome of the conflict in region 1 depends on whether

$$\theta_1^c > \theta_1^e + \sigma,$$

where  $\sigma$  is a random variable drawn from a cumulative distribution  $H(\cdot)$ . If this inequality holds, then the citizen side wins and determines the taxes, transfers and the public good provision level. If it does not, the elite side wins and makes all the policy decisions.

Given this specification, the probability that citizens in region 1 win this conflict is simply

$$H(\theta_1^c - \theta^e).$$

Let us next turn to the case in which still  $s = 0$  and at least one of  $N^{z^c=1}$  and  $N^{z^e=1}$  is not a singleton, so that conflict is not purely local. Then, analogously, the outcome of the conflict depends on whether

$$\sum_{i \in N^{z^c=1}} \theta_i^c > |N^{z^e=1}| \theta^e + \sigma. \tag{2}$$

If this inequality holds, the coalition of citizens wins, and otherwise the elite coalition wins, and makes the policy choices. Intuitively, the left-hand side involves the investments of all citizen groups that are in a coalition with a strong partner, and thus able to take part in a conflict. The right-hand side involves the strength of all elite groups that have joined the coalition involving region 1. The right-hand side thus reflects the fact that all their resources are pooled. Notice that this expression also applies when  $N^{z^c=1}$  is not equal to  $N^{z^e=1}$ . If some region  $j \in N^{z^c=1}$  but  $j \notin N^{z^e=1}$ , it means that they are contributing to the elite side of this fight, and since  $j \notin N^{z^e=1}$ , citizens from this region are not contributing to the citizen side. The cost to the elite from region  $j$  is that if the citizen side wins, they will have also lost, whereas if they had not joined this coalition, because their citizens are weak, they would have never lost the conflict. Conversely, if  $j \in N^{z^c=1}$  but  $j \notin N^{z^e=1}$ , then the outcome of the conflict in region  $j$  is determined depending on whether

$$\sum_{i \in N^{z^c=1}} \theta_i^c > \theta^e + \sigma,$$

implying that the elite in this region are facing the full strength of the citizen coalition  $N^{z^c=1}$ .

<sup>11</sup> There is no interesting possibility that  $N^{z^e=1}$  is a singleton but not equal to  $\{1\}$ , since in this case the citizens are not organized, and thus there is no need to make any concessions.

Finally, consider the case in which  $s = 1$ , so that there is state centralization, and the citizen coalition is given by  $\mathcal{N}^{z^c=1}$ . Because state centralization has already pooled all elite resources, the outcome of the conflict now depends on whether

$$\sum_{i \in \mathcal{N}^{z^c=1}} \theta_i^c > N\theta^e + \sigma.$$

In all of this, the conflict always destroys a fraction  $1 - \alpha \in (0, 1)$  of the total output, representing the fact that conflict is costly. Hence, the party that wins the conflict will have access to a regional output in the amount of  $\alpha Y$ . This cost can be avoided if the elite in question choose to make an offer  $(T_b, G_i)$  that the citizen side prefers to fighting, and thus chooses  $f_i^e = 0$ .

Finally, in what follows we will also impose:

**Assumption 2.** The density of the distribution function  $H, h$ , exists over its entire support  $\mathcal{H} \subset [-N\theta^e, \bar{\theta})$  where  $\bar{\theta} > 0$ , is continuously differentiable and is nonincreasing. Moreover,  $h(-N\theta^e)\alpha Y > \Gamma'(0)$ .

This assumption is useful for several reasons. First, it ensures that the second-order condition in the conflict choice of citizens is satisfied. Second it guarantees that the density of the distribution function  $H$ , which shapes the marginal incentives of citizens in their investment decisions, is well-defined and positive over the range in which these investments will take place ( $\bar{\theta}$  is an arbitrary positive constant, making sure that the support of the distribution does not stop exactly at 0). Third, it also ensures that starting at zero conflict capacity, citizens have an incentive for investment in this capacity. This assumption is the weaker version of the oft-imposed requirement that  $\Gamma'(0) = 0$ .

### 2.3. Timing of events and equilibrium

To summarize, the timing of events is as follows.

1. The elites decide whether to centralize the state (i.e., choose between  $s = 0$  and  $s = 1$ ). Note that at this stage, all regional elites have the same preferences over state centralization.
2. It becomes common knowledge in which region citizens are strong. Suppose, without loss of any generality that this is region 1. Then citizens in region 1 decide whether to form a coalition with other regions (i.e., choose between  $Z^c = 0$  and  $Z^c = 1$ ). If  $Z^c = 0$ , then there is no coalition of citizens from different regions. If  $Z^c = 1$ , then other regions decide whether to join the coalition of the strong citizens from region 1 (i.e., they choose  $z_i^c = 0$  or  $z_i^c = 1$  for  $i = 2, \dots, N$ ). In region 1 and those in  $i \in \mathcal{N}^{z^c=1}$  (or equivalently, those where  $z_i^c = 1$ ) citizens choose  $\theta_i^c \geq 0$ .
3. Then elites from different regions decide whether to join in a coalition with the elite from region 1, which are the ones facing the strong citizens (i.e., they decide  $z_i^e = 0$  or  $z_i^e = 1$  for  $i = 2, \dots, N$ ). Then:
  - (a) In regions  $i \notin \mathcal{N}^{z^c=1} \cup \mathcal{N}^{z^e=1}$  (or equivalently those with  $i \geq 2$  and  $z_i^c = z_i^e = 0$ ) political power is not contested and the elites decide the policy vector  $(T_b, G_i)$ .
  - (b) Elites that are in  $\mathcal{N}^{z^c=1}$  jointly decide what offer  $(T_i, G_i)_{i \in \mathcal{N}^{z^c=1}}$  to make to the citizens they are facing, and those in  $\mathcal{N}^{z^c=1} \setminus \mathcal{N}^{z^e=1}$  individually make such offers.
4. If the state is centralized ( $s = 1$ ), then elites from different regions will have already formed their grand coalition, i.e.,  $\mathcal{N}^{z^e=1} = \mathcal{N}$ . Then, all of the elites jointly decide what offer  $(T_i, G_i)_{i \in \mathcal{N}^{z^e=1}}$  to make to the citizens they are facing.
5. Following these offers, citizens decide whether to accept the offers they have received or not ( $f_{\mathcal{N}^{z^c=1}}^e \in \{0, 1\}$  or  $f_i^e \in \{0, 1\}$ ). If the offer is accepted, it is implemented. Otherwise, there is fighting and whether the citizens or the elite win is determined according to (2), and the winner sets the policies with the value of output reduced to  $\alpha Y$ .
6. Policies are implemented, and all payoffs are realized.

In what follows, we focus on pure-strategy subgame perfect equilibria. This detailed timing of events also specifies citizen and elite strategies, and a subgame perfect equilibrium is defined, in the usual fashion, as a strategy profile in which all actions are best responses to other strategies in all histories. When this will cause no confusion, we refer to pure-strategy subgame perfect equilibria simply as “equilibria”.

**Remark 2.** The timing of events also clarifies that there are two different ways in which the elites can “centralize the state” (form their grand coalition and coordinate their actions). The first is by choosing  $s = 1$  in stage 1, and the second one is by choosing  $\mathcal{N}^{z^c=1} = \mathcal{N}$  in stage 3. As we have specified the payoffs, these two options are entirely equivalent. It is straightforward, but cumbersome, to introduce a slight cost advantage for the first option, so that elites explicitly choose  $s = 1$  when this is in their interest rather than wait for stage 4. In what follows, we simplify the discussion by assuming that when state centralization is in their interest, the elites will do so by setting  $s = 1$ .

### 3. Equilibrium

Pure-strategy subgame perfect equilibria are characterized by backward induction. We start by the policy offer of elites under threat of conflict and the response of citizens.

**Lemma 1.** Regardless of whether  $s = 0$  or  $s = 1$ , any equilibrium always involves  $f_{\mathcal{N}^{z^c=1}}^e = 0$ , i.e., there will always be an offer from the elite



that induces no fighting. Moreover, this offer will give citizens exactly the same utility as they would obtain with fighting.

**Proof.** See the Appendix. □

The intuition for this is simple. Since  $\alpha < 1$ , conflict is costly, and the elites can always benefit by offering the policy mix that makes citizens as well off as they would be with conflict. Moreover, since the elite have the possibility to make such an offer, they will never propose a policy mix that gives citizens strictly greater utility than the latter could obtain by fighting. This last observation also implies that in the previous stages of the game in our analysis of the decisions of citizens, we could always use the utility that they would obtain under fighting.

The next question is what policy mix the elites will use, what coalitions will form along the equilibrium path, and whether the elites will choose state centralization. To investigate these issues, we first characterize the equilibria in subgames starting first without state centralization, and then with state centralization.

### 3.1. Equilibrium without a centralized state ( $s = 0$ )

Suppose that the elites have decided not to form a centralized state, designated by  $s = 0$ .

#### 3.1.1. The escalation effect

Our next result formalizes the escalation effect in a society without state centralization. It shows that when there has been no state centralization, the coalition formation of elites will mimic that of citizens. Throughout, we continue to suppose, without loss of any generality, that region 1 is the one where citizens are strong.

**Lemma 2.** Suppose  $s = 0$  (there has been no state centralization) and citizens have formed a coalition  $\mathcal{N}^{z^c=1} (\ni 1)$ . Then

$$\mathcal{N}^{z^e=1} = \mathcal{N}^{z^c=1}.$$

**Proof.** See the Appendix. □

Intuitively, no elite in a region where the citizens have not joined the coalition  $\mathcal{N}^{z^c=1}$  would want to join the coalition  $\mathcal{N}^{z^e=1}$ , since they are facing weak, unorganized citizens that cannot make any demands, but if they join the coalition, this will force them to make concessions or be included in the fight with positive probability of losing (because at least some other members of the coalition are facing organized citizens). Conversely, elites in the regions where citizens have joined the coalition  $\mathcal{N}^{z^c=1}$  will be facing organized demands, and are better off pooling their resources with other elites.

The anticipation of this behavior highlights the *escalation effect* mentioned in the Introduction: when citizens in region 1 decide to form a coalition with citizens from other regions, they will escalate the conflict, inducing other elite groups to join the fight as well.<sup>12</sup>

#### 3.1.2. Choice of conflict capacity

Suppose now that a coalition  $\mathcal{N}^{z^c=1}$  of citizens has formed. How will they choose their conflict capacity? First, we know from Lemma 2 that  $\mathcal{N}^{z^e=1} = \mathcal{N}^{z^c=1}$ . Next recall that even though the group of citizens will make their fighting decisions jointly, the level of conflict capacity is the purview of each region.<sup>13</sup> Hence, it will be the solution to a maximization problem in which each group  $i$  of citizens in the coalition  $\mathcal{N}^{z^c=1}$  chooses  $\theta_i^c$  recognizing that they will be facing an identical coalition of elites. Setting  $|\mathcal{N}^{z^c=1}| = n$ , this maximization problem for each  $i \in \mathcal{N}^{z^c=1}$  is:

$$\max_{\theta_i^c \geq 0} H \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - n\theta^e \right) \Phi(n)\alpha Y - \Gamma(\theta_i^c). \tag{3}$$

Intuitively,  $H \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - n\theta^e \right)$  is the probability that the citizens will win the conflict — since their total conflict capacity will be  $\sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c$  and the exogenous conflict capacity of the elite they are facing is  $n\theta^e$ . If they lose in the conflict, then the regional elites choose the policies and naturally set zero taxes, yielding zero utility to citizens. If the citizens succeed, they can tax the entire income that is not destroyed in conflict,  $\alpha Y$ , and they can use this either for the direct transfers or public good investments, and the term  $\Phi(n)$

<sup>12</sup> As already noted in the Introduction, the escalation effect does not always harm the strong citizens (citizens from region 1). For instance, if the strength of elites,  $\theta^e$ , is small and the level of investment of citizens,  $\theta^c$ , is large, the likelihood that citizens prevail in the conflict when both sides have formed their grand coalitions (given by the probability that  $|\mathcal{N}|\theta^c > |\mathcal{N}|\theta^e + \sigma$ ) will be greater than the likelihood that citizens from region 1 win against elites from region 1 (given by the probability that  $\theta_1^c > \theta^e + \sigma$ ). However, as we will see below, the escalation effect will always harm citizens from region 1 at the *margin* when they are indifferent between engaging in conflict by themselves and forming a larger coalition (e.g., their grand coalition). This is because for them to be indifferent in this fashion, we have to be in the case where  $\theta^e$  is relatively large compared to  $\theta^c$ .

<sup>13</sup> This assumption is made so as to ensure that forming a coalition does not automatically increase their investments in conflict capacity by removing the free-rider effect (which thus prevents us from mixing two potentially distinct benefits of forming a larger coalition, the first coming from the political agenda effect, and the second one from the wrecked coordination to remove the free-rider effect). In the Appendix we characterize the equilibrium of the model when citizens solve the free-rider effect in their investment in conflict capacity, and show that all of our results apply identically in this case (though the exact thresholds are different).

captures the marginal utility of these funds when they are optimally used. Finally,  $\Gamma(\theta_i^c)$  is the cost that this group of citizens faces from their investments.

The first-order condition for this problem is (for each  $i \in \mathcal{N}^{z^c=1}$ )

$$h \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - n\theta^e \right) \Phi(n)\alpha Y - \Gamma'(\theta_i^c) = 0, \tag{4}$$

with the second-order condition

$$h' \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - n\theta^e \right) \Phi(n)\alpha Y - \Gamma''(\theta_i^c) < 0.$$

Assumption 2 ensures that the first-order condition (4) will always hold, thus removing the need to write this in complementary slackness form, and also that there will be a unique solution where the second-order condition holds (since  $h' \leq 0$ ).

One important implication of (4) is that citizens from all regions will choose the same investment in conflict capacity, and this will depend only on  $n$  (and not on the exact identity of the regions in  $\mathcal{N}^{z^c=1}$ ).<sup>14</sup> We denote this investment level by  $\theta^*(|\mathcal{N}^{z^c=1}|, |\mathcal{N}^{z^e=1}|, \Phi(|\mathcal{N}^{z^c=1}|))$ , where the first argument is the conditioning on the size of the coalition of citizens, the second argument is the size of the coalition of elites the citizens will be facing, and the term  $\Phi(|\mathcal{N}^{z^c=1}|)$  highlights the other major effect discussed in the Introduction, *the political agenda effect*. The presence of this term, and thus the political agenda effect, both directly increases the utility of citizens from conflict in (3) and raises their level of investment in conflict capacity in (4). Since we are in the case of no state centralization and  $|\mathcal{N}^{z^e=1}| = |\mathcal{N}^{z^c=1}| = n$  (from Lemma 2), the equilibrium level of conflict capacity can be simply denoted  $\theta^*(n, n, \Phi(n))$ .

Indeed, because  $\Gamma'' > 0$  everywhere, the level of investment in conflict capacity is always (strictly) increasing in whatever increases the marginal utility of additional investments, given by  $h(\sum_{i \in \mathcal{N}^{z^c=1}} \theta_i^c - n\theta^e)\Phi(n)\alpha Y$ . The political agenda effect then follows straightforwardly from this observation, since

$$\partial \theta^*(n, n, \Phi(n)) / \partial \Phi(n) > 0,$$

and  $d\Phi(n)/dn \geq 0$ . Intuitively, a higher  $\Phi(n)$  implies that the marginal utility of the funds that citizens can capture following a successful conflict is greater, and this will encourage them to invest more in conflict. In turn, larger coalitions of citizens can use funds more productively to provide public goods, thus explaining why  $\Phi(n)$  is increasing in  $n$ .

More specifically, substituting these equilibrium conflict investments in the utility functions of citizens, we can observe that the utility of citizens from each region  $i \in \mathcal{N}^{z^c=1}$  under a non-centralized state ( $s = 0$ ) is

$$U_i^{c*}[n|s = 0] = \alpha H(n\theta^*(n, n, \Phi(n)) - n\theta^e)\Phi(n)Y - \Gamma(\theta^*(n, n, \Phi(n))). \tag{5}$$

From Lemma 1, the elite will make an offer to citizens that just convinces them not to fight, and at this point they have already paid the cost of investments in conflict capacity. Hence, the utility of citizens from each region  $i \in \mathcal{N}^{z^c=1}$  entering conflict will be  $H(n\theta^*(n, n, \Phi(n)) - n\theta^e)\Phi(n)\alpha Y$ . With the same logic that citizens can use their funds most effectively by providing public goods, elites in the coalition  $\mathcal{N}^{z^e=1} = \mathcal{N}^{z^c=1}$  can also deliver this utility most effectively by providing public goods, since one unit of the consumption good invested in the public good equally across each region in  $\mathcal{N}^{z^c=1}$  yields the utility of  $\Phi(n)$  to each group of citizens. Hence, the cost of delivering a utility of  $H(n\theta^*(n, n, \Phi(n)) - n\theta^e)\Phi(n)\alpha Y$  to citizens in this coalition is  $H(n\theta^*(n, n, \Phi(n)) - n\theta^e)\alpha Y$ . Thus the utility of elites in  $i \in \mathcal{N}^{z^c=1}$  can be written as

$$U_i^{e*}[n|s = 0] = [1 - \alpha H(n\theta^*(n, n, \Phi(n)) - n\theta^e)]Y. \tag{6}$$

Because of the political agenda effect,  $U_i^{c*}[n|s = 0]$  is increasing in  $\Phi(n)$ , while  $U_i^{e*}[n|s = 0]$  is decreasing in  $\Phi(n)$ . The latter is true despite the fact that, when the citizens have formed a coalition of size  $n$ , each dollar that the elite decides to redistribute is also worth more by exactly the same amount,  $\Phi(n)$ , since in this case a higher  $\Phi(n)$  also raises the investment of citizens in their conflict capacity.

Second, because  $h' \leq 0$ , the marginal utility of funds is decreasing in the investments of other citizens in the coalition and in the size of the coalition, making investments of different groups of citizens strategic substitutes. For the same reason, the marginal utility of funds is increasing in the size of the coalition of elites, creating a force towards greater investments when citizens are facing a larger coalition of elites.

### 3.1.3. Coalition decision of citizens

Let us now consider the coalition decision of citizens. If  $Z^c = 1$ , then clearly  $z_i^c = 1$  for all  $i = 2, \dots, N$ . This is because when they do not join a coalition with citizens from region 1, these unorganized groups of citizens cannot contest power and thus receive zero transfers in public goods, whereas once they do so, they will receive positive transfers in the next stage of the game.

This then implies that if  $Z^c = 1$ , the grand coalition of all citizens will form. Consequently, the choice for citizens from region 1 is

<sup>14</sup> This applies to citizens from region 1 as well, since, after joining the coalition, the problem facing all citizens, including those from region 1, are identical.

to choose  $Z^c = 0$  and act by themselves, or initiate the formation of this grand coalition.<sup>15</sup>

Suppose first that  $\mathcal{N}^{z^c=1} = \{1\}$ . Then with the same notation as above and noting that in this case the size of the coalition is 1 and also from Assumption 1)  $\Phi(1) = 1$ , (5) becomes

$$U_1^{e*}[1|s = 0] = \alpha H(\theta^*(1, 1, 1) - \theta^e)Y - \Gamma(\theta^*(1, 1, 1)). \tag{7}$$

For future reference, we also write the utility of elites from region 1 in this case, which follows readily from (6):

$$U_1^{e*}[1|s = 0] = [1 - \alpha H(\theta^*(1, 1, 1) - \theta^e)]Y. \tag{8}$$

Note also that in this case, the most efficient way of transferring resources for the elite is a direct transfer, so there will be no investment in public goods. That is,  $T_1 = \alpha H(\theta^*(1, 1, 1) - \theta^e)Y$  and  $G_1 = 0$ .

Suppose next that  $\mathcal{N}^{z^c=1} = \mathcal{N}$ . Then because the equilibrium is symmetric and thus citizens from different regions will choose the same investment level, (5) becomes

$$U^{c*}[N|s = 0] = \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^e)\Phi(N)Y - \Gamma(\theta^*(N, N, \Phi(N))), \tag{9}$$

and similarly, (6) becomes

$$U^{e*}[N|s = 0] = [1 - \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^e)]Y \tag{10}$$

Since from Assumption 1,  $\Phi(N) > 1$ , we also know that all transfers from elites to citizens in this case will take the form of public good provision. That is,  $T_i = 0$  and  $G_i = \mu\alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^e)Y$  for all  $i$ .

We can conclude that citizens from region 1 will prefer to form the grand coalition of citizens when their utility from the grand coalition, (9), is strictly greater than their utility from standing alone, (7). The former, (9), is increasing in  $\Phi(N)$ ,<sup>16</sup> while the latter, (7), does not depend on  $\Phi(N)$ . Therefore, there exists a value of  $\Phi(N)$ ,  $\Phi^*(N)$ , such that

$$\begin{aligned} H(N\theta^*(N, N, \Phi^*(N)) - N\theta^e)\Phi^*(N)\alpha Y - \Gamma(\theta^*(N, N, \Phi^*(N))) \\ = H(\theta^*(1, 1, 1) - \theta^e)\alpha Y - \Gamma(\theta^*(1, 1, 1)). \end{aligned} \tag{11}$$

Then whenever  $\Phi(N) > \Phi^*(N)$ , citizens from region 1 (the strong citizens) prefer a grand coalition, and whenever  $\Phi(N) < \Phi^*(N)$ , they prefer to act by themselves. This conclusion is intuitive. Forming a coalition with citizens from other regions is costly because of the escalation effect: the fight will escalate because, in response, the grand coalition of elites will form and fight against them (this is captured by the fact that  $\theta^e$  is now multiplied by  $N$  on the left-hand side). The benefit, on the other hand, is related to the political agenda effect as well as to the fact that now all citizens are coordinated (captured by the fact that  $\theta^*(N, N, \Phi(N))$  is multiplied by  $N$ ). The latter effect corresponds to the more effective use of funds in case of victory in the conflict, and exhibits itself in the presence of the term  $\Phi(N)$  multiplying the probability of success on the left-hand side, and the greater level of investment in conflict capacity of all citizens (i.e., the fact that  $\theta^*(N, N, \Phi(N)) > \theta^*(N, N, 1)$ ).

Another important point is worth noting: equation (11) is purely from the viewpoint of citizens in region 1, trading off the escalation effect against the coordination of all citizens and the political agenda effect. However, when citizens from region 1 form a coalition with other citizens, they create a positive externality on these citizens, who would have otherwise not contested power and obtained zero transfers and public goods, and a commensurate negative externality on elites from other regions, which now will have to make transfers to their citizens. The ex ante expectation of this externality plays an important role in interpreting our results below.

The coalition formation decision of citizens, and the resulting utility levels for citizens and elites, are summarized in the next proposition.

**Proposition 1.** *Suppose the elites have not formed a centralized state (i.e.,  $s = 0$ ), and suppose without loss of any generality that it is citizens from region 1 that are strong. Then there exists  $\Phi^*(N)$  such that the following are true:*

1. *If  $\Phi(N) < \Phi^*(N)$ , then citizens from region 1 choose  $Z^c = 0$  and act independently. There is no redistribution in regions  $i = 2, \dots, N$ , and in region 1, there is redistribution in the form of direct transfers to convince citizens in this region not to fight. That is,  $T_1 = \alpha H(\theta^*(1, 1, 1) - \theta^e)Y$  and  $G_1 = 0$ .*
2. *If  $\Phi(N) > \Phi^*(N)$ , then citizens from region 1 choose  $Z^c = 1$  and all of the regions choose  $z_i^c = 1$ , joining the coalition of citizens from region 1. In this case, there will be redistribution in all regions in the form of public good provision. That is,  $T_i = 0$  and  $G_i = \mu\alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^e)Y$  for all  $i$ .*

**Proof.** The existence of the threshold  $\Phi^*(N)$  follows from (11). Then by the definition of  $\Phi^*(N)$ , when  $\Phi(N) < \Phi^*(N)$ , the right-hand side of (11) is greater and thus citizens from region 1 prefer  $Z^c = 0$ , and the rest of part 1 follows straightforwardly. When

<sup>15</sup> As already noted, in the Appendix we discuss the case in which citizens from region 1 can restrict the coalition to a certain number of regions, thus inducing the formation of a smaller coalition, and show that when  $H(\cdot)$  is uniform and  $\Gamma(\cdot)$  is quadratic, they will never choose to do this. We also establish that when  $N$  is sufficiently large, even if citizens choose to form a smaller coalition, the elites will opt for a non-centralized state.

<sup>16</sup> The derivative is given by:  $H(N\theta^*(N, N, \Phi(N)) - N\theta^e)\alpha Y + \frac{\partial \theta^*(N, N, \Phi(N))}{\partial \Phi(N)} [Nh(N\theta^*(N, N, \Phi(N)) - N\theta^e)\Phi(N)\alpha Y - \Gamma'(\theta^*(N, N, \Phi(N)))]$ . The first term is clearly positive. We can further see that the second term is positive by the following argument. First,  $\frac{\partial \theta^*(N, N, \Phi(N))}{\partial \Phi(N)} > 0$  with a straightforward application of the implicit function theorem. Second, the term in square brackets in the second line is also strictly positive, since the first-order condition (4) implies that  $h(N\theta^*(N, N, \Phi(N)) - N\theta^e)\Phi(N)\alpha Y = \Gamma'(\theta^*(N, N, \Phi(N)))$ , and thus this term is equal to  $(N - 1)h(N\theta^*(N, N, \Phi(N)) - N\theta^e)\Phi(N)\alpha Y > 0$ .

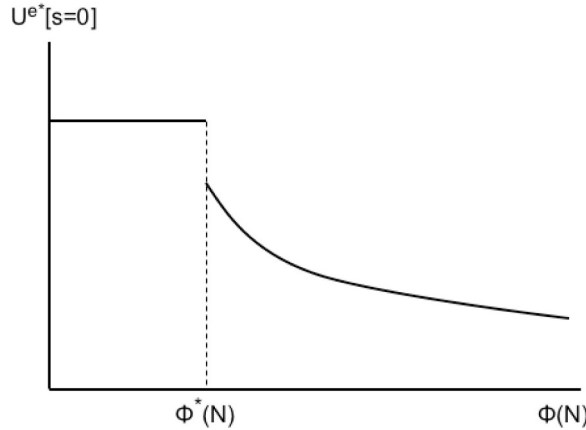


Fig. 1.

$\Phi(N) > \Phi^*(N)$ , the left-hand side of (11) is greater, and citizens from region 1 prefer  $Z^c = 1$ . The rest of part 2 once again follows readily.  $\square$

An important point, related to our discussion in the Introduction, is that when  $\Phi(N) < \Phi^*(N)$  and conflict is local, elites placate the citizens by making direct transfers to them instead of providing public goods. Instead, when  $\Phi(N) > \Phi^*(N)$  and conflict is national, there will be public good provision instead of direct transfers. This emphasizes the link between the nature of conflict and demands from citizens and whether the responses from elites are “patrimonial”.

We can also readily determine the expected utility of elites before the identity of the citizens that are strong is determined. When  $\Phi(N) < \Phi^*(N)$ , there is only  $1/N$  probability that they will have to redistribute income to their citizens as specified in Proposition 1, whereas when  $\Phi(N) > \Phi^*(N)$ , all elites will necessarily make transfers as specified in Proposition 1. Therefore, the following corollary to Proposition 1 follows immediately (proof in the text).

**Corollary 1.** *When  $\Phi(N) < \Phi^*(N)$ , the expected utility of elites is*

$$U^{e*}[s = 0] = \left[ 1 - \frac{\alpha}{N} H(\theta^*(1, 1, 1) - \theta^e) \right] Y. \tag{12}$$

*When  $\Phi(N) > \Phi^*(N)$ , the expected utility of elites is*

$$U^{e*}[s = 0] = [1 - \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^e)] Y. \tag{13}$$

These expressions make it clear that, as shown in Figure 1, when  $\Phi(N)$  crosses the threshold of  $\Phi^*(N)$ , there is a downward jump in the utility of elites due to the change in the equilibrium coalition of citizens.<sup>17</sup> This discontinuous behavior at the threshold  $\Phi^*(N)$  reflects the political agenda effect, and highlights the advantage of a non-centralized state for the elites when the marginal utility of funds for citizens is not too high (i.e., is less than  $\Phi^*(N)$ ). It is this advantage (in the political agenda effect), as we will see next, that will induce elites to opt for a non-centralized state for a range of values of  $\Phi(N)$ .

### 3.2. Equilibrium under a centralized state

We next present the analogous analysis for the case in which there is a centralized state, i.e.,  $s = 1$ .

#### 3.2.1. Choice of conflict capacity

The choice of conflict capacity is very similar to that under the non-centralized state, except that now the elite have already committed to nationally-coordinated action. Therefore, when the coalition of citizens is  $\mathcal{N}^{z^c=1}$ , the optimization problem for citizens from region  $i \in \mathcal{N}^{z^c=1}$  becomes

$$\max_{\theta_i^c \geq 0} H \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - N\theta^e \right) \Phi(n) \alpha Y - \Gamma(\theta_i^c),$$

where  $|\mathcal{N}^{z^c=1}| = n$ . Here we have used the fact that because of state centralization, the baseline includes a greater amount of elite resources being deployed against the citizens. The first-order conditions are then identical to (4), with the  $n\theta^e$  term replaced by  $N\theta^e$ ,

<sup>17</sup> That the jump is downward follows from the same argument as the one we use in step 1 of the proof of Proposition 3.

and we do not write them out to conserve space. One important implication is that the escalation effect which played an important role in our analysis without state centralization is no longer present — the elites have already banded together and there will be no further escalation of conflict on the elite side even if citizens form a non-local coalition.

3.2.2. The coalition formation decision of citizens

The coalition decision of citizens can be analyzed similarly to the one without state centralization. First, it is again straightforward to see that when citizens from region 1, who we have again set as the ones that are strong without loss of any generality, choose  $Z^c = 1$ , then citizens from all other regions will choose  $z_i^c = 1$ . Thus the choice for citizens from region 1 is once more simply between  $N^{z^c=1} = \{1\}$  and  $N^{z^c=1} = N$ . When they choose  $N^{z^c=1} = \{1\}$ , we again have  $\Phi(1) = 1$ , and thus their overall utility under state centralization ( $s = 1$ ) will be

$$U_1^{c*}[1|s = 1] = \alpha H(\theta^*(1, N, 1) - N\theta^e)Y - \Gamma(\theta^*(1, N, 1)). \tag{14}$$

Conversely, when they choose  $N^{z^c=1} = N$ , their overall utility under state centralization will be

$$U_1^{c*}[N|s = 1] = \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^e)\Phi(N)Y - \Gamma(\theta^*(N, N, \Phi(N))). \tag{15}$$

The comparison of these two expressions boils down to

$$\begin{aligned} &H(N\theta^*(N, N, \Phi(N)) - N\theta^e)\Phi(N)\alpha Y - \Gamma(\theta^*(N, N, \Phi(N))) \\ &> H(\theta^*(1, N, 1) - N\theta^e)\alpha Y - \Gamma(\theta^*(1, N, 1)). \end{aligned} \tag{16}$$

When evaluated at  $\Phi(N) = \Phi^*(N)$ , the left-hand side of (16) is identical to the left-hand side of (11), while its right-hand side is strictly smaller than the right-hand side of (11) because of the presence of  $N$  multiplying  $\theta^e$  inside the  $H$  function. Intuitively, this reflects the absence of the escalation effect mentioned above — citizens can now make their coalition formation decision without worrying about how this will escalate the conflict by inducing the elites to form a coalition, and therefore they are more likely to form such a coalition. These observations also establish the next proposition.

**Proposition 2.** *Suppose the elites have formed a centralized state (i.e.,  $s = 1$ ), and suppose without loss of any generality that citizens from region 1 are strong. Then citizens from region 1 choose  $Z^c = 1$  and all the regions choose  $z_i^c = 1$ , joining the coalition of citizens from region 1. There will be redistribution in all regions in the form of public good provision. That is,  $T_i = 0$  and  $G_i = \mu\alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^e)Y$  for all  $i$ .*

**Proof.** The proof is straightforward by observing that (16) is always satisfied. To see this note that since  $\Phi(N) \geq 1$  and the left-hand side of (16) is increasing in  $\Phi(N)$  while the right-hand side is independent of  $\Phi(N)$ , a stronger (more demanding) condition than (16) is that

$$H(N\theta^*(N, N, 1) - N\theta^e)\alpha Y - \Gamma(\theta^*(N, N, 1)) > H(\theta^*(1, N, 1) - N\theta^e)\alpha Y - \Gamma(\theta^*(1, N, 1)). \tag{17}$$

To see that this must always hold, we first prove that

$$H(N\theta^*(N, N, 1) - N\theta^e)\alpha Y > H(\theta^*(1, N, 1) - N\theta^e)\alpha Y. \tag{18}$$

Suppose, to obtain a contradiction, that this is not the case. From (4), we have

$$\begin{aligned} h(N\theta^*(N, N, 1) - N\theta^e)\alpha Y &= \Gamma'(\theta^*(N, N, 1)) \\ h(\theta^*(1, N, 1) - N\theta^e)\alpha Y &= \Gamma'(\theta^*(1, N, 1)). \end{aligned}$$

But since  $h$  is nonincreasing from Assumption 2,

$$h(N\theta^*(N, N, 1) - N\theta^e) \geq h(\theta^*(1, N, 1) - N\theta^e).$$

Then, given that  $\Gamma' > 0$ ,  $\theta^*(N, N, 1) > \theta^*(1, N, 1)$ . This in turn implies that  $N\theta^*(N, N, 1) - N\theta^e > \theta^*(1, N, 1) - N\theta^e$ , and thus

$$H(N\theta^*(N, N, 1) - N\theta^e)\alpha Y > H(\theta^*(1, N, 1) - N\theta^e)\alpha Y,$$

yielding a contradiction.

Next, since (18) holds, we must have from (4) and Assumption 2 that  $\Gamma'(\theta^*(N, N, 1)) < \Gamma'(\theta^*(1, N, 1))$ , and hence that  $\Gamma(\theta^*(N, N, 1)) < \Gamma(\theta^*(1, N, 1))$ , establishing that (17), and thus (16), is always satisfied.  $\square$

Intuitively, because the escalation effect is entirely absent, citizens simply gain additional strength by joining together without any cost of doing so. This means that when elites form a centralized state, citizens will also respond by forming their grand coalition. This result thus highlights the cost of state centralization from the viewpoint of elites. It implies that citizens will necessarily form a coalition in their potential conflict against the elites.<sup>18</sup>

<sup>18</sup> The result that citizens will form their grand coalition under a centralized state is due to the assumption that the escalation effects disappear entirely under state centralization. If we relax this, for example, by assuming that the centralized state is equivalent to elites pooling some fraction of their resources (thus giving them a strength  $K\theta^e$ , where  $K \in (1, N)$ ), and their remaining power will be pooled in response to citizens joining together in conflict, then there will exist another threshold,  $\bar{\Phi}(N) < \Phi^*(N)$  such that the citizens will form their grand coalition under state centralization only if  $\Phi(N) > \bar{\Phi}(N)$ .



The next corollary once again characterizes the expected utility of elites.

**Corollary 2.** *When there is state centralization ( $s = 1$ ), the equilibrium utility of elites is given by*

$$U^{e*}[s = 1] = [1 - \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^e)]Y. \tag{19}$$

#### 4. Equilibrium state centralization

In this section, we study the state centralization decision of elites by combining the results from Corollaries 1 and 2. We will first provide a sharp result under Assumption 2, which ensures that  $h$  is nonincreasing (i.e.,  $h'(\cdot) \leq 0$ ), and then in the next subsection we will generalize this result to the case in which  $h$  is single-peaked, which will reveal an additional strategic interplay.

##### 4.1. Main result

**Proposition 3.** *Let  $\Phi^*(N)$  be as defined in Proposition 1, and suppose that Assumptions 1 and 2 hold. Then there exists a  $\Phi_*(N)$  such that we have the following characterization of the unique equilibrium:*

1. *If  $\Phi(N) < \Phi_*(N)$ , then the elites choose state centralization ( $s = 1$ ).*
2. *If  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$ , then the elites choose a non-centralized state ( $s = 0$ ).*
3. *If  $\Phi^*(N) < \Phi(N)$ , then the elites choose state centralization ( $s = 1$ ).<sup>19</sup>*

**Proof.** We will proceed in several steps.

*Step 1:* We first prove that for  $\Phi^*(N) \geq 1$

$$H(N\theta^*(N, N, \Phi^*(N)) - N\theta^e)\Phi^*(N)\alpha Y \geq H(\theta^*(1, 1, 1) - \theta^e)\alpha Y. \tag{20}$$

Suppose, to obtain a contradiction, that this is not the case. From (4), we have

$$\begin{aligned} h(N\theta^*(N, N, \Phi^*(N)) - N\theta^e)\Phi^*(N)\alpha Y &= \Gamma'(\theta^*(N, N, \Phi^*(N))) \\ h(\theta^*(1, 1, 1) - \theta^e)\alpha Y &= \Gamma'(\theta^*(1, 1, 1)). \end{aligned} \tag{21}$$

But since  $h$  is nonincreasing from Assumption 2,

$$h(\theta^*(1, 1, 1) - \theta^e) \leq h(N\theta^*(N, N, \Phi^*(N)) - N\theta^e)\Phi^*(N).$$

But then, given that  $\Gamma' > 0$ ,  $\theta^*(N, N, \Phi^*(N)) \geq \theta^*(1, 1, 1)$ . This immediately implies that,  $\Gamma(\theta^*(N, N, \Phi^*(N))) \geq \Gamma(\theta^*(1, 1, 1))$ , and thus

$$H(N\theta^*(N, N, \Phi^*(N)) - N\theta^e)\Phi^*(N)\alpha Y \geq H(\theta^*(1, 1, 1) - \theta^e)\alpha Y,$$

yielding a contradiction, establishing step 1.

*Step 2:* Define  $\Phi_*(N)$  such that

$$H(N\theta^*(N, N, \Phi_*(N)) - N\theta^e) = \frac{1}{N}H(\theta^*(1, 1, 1) - \theta^e).$$

Since  $\theta^*(N, N, \Phi_*(N))$  is increasing in  $\Phi_*(N)$  (from (4)), for all  $\Phi(N) > \Phi_*(N)$ , the elite prefer non-state centralization, and for all  $\Phi(N) < \Phi_*(N)$ , they prefer state centralization.

*Step 3:* We next prove that in the relevant region where  $\Phi^*(N) \geq 1$ , we have  $\Phi_*(N) < \Phi^*(N)$  (if  $\Phi^*(N) \leq 1$ , then we could not have  $\Phi(N) > \Phi^*(N)$ ).

Recall from step 1 that  $H(N\theta^*(N, N, \Phi^*(N)) - N\theta^e)\Phi^*(N) \geq H(\theta^*(1, 1, 1) - \theta^e)$ . Dividing both sides by  $N$ , we have

$$H(N\theta^*(N, N, \Phi^*(N)) - N\theta^e)\frac{\Phi^*(N)}{N} \geq \frac{1}{N}H(\theta^*(1, 1, 1) - \theta^e).$$

Since from (1) and Assumption 1,  $\Phi^*(N) = \Phi(N) = (N - (N - 1)\lambda)\mu < N$ , this implies

$$H(N\theta^*(N, N, \Phi^*(N)) - N\theta^e) > \frac{1}{N}H(\theta^*(1, 1, 1) - \theta^e),$$

which yields that for  $\Phi(N) = \Phi^*(N)$ , (12) is strictly greater than (19). Since  $\Phi(N) > \Phi_*(N)$  and (12) is strictly greater than (19), we must have that  $\Phi^*(N) > \Phi_*(N)$ .

We now proceed to prove the three claims in the proposition.

**Part 1:** If  $\Phi(N) < \Phi_*(N)$ , (12) is strictly less than (19), and thus the elites prefer a centralized state with citizens forming their

<sup>19</sup>Throughout we simplify the statement of the propositions by omitting the case where  $\Phi(N) = \Phi^*(N)$ , which involves the citizens being indifferently between joining in a coalition and not.

grand coalition to a non-centralized state where there is no coalition formation among citizens. Therefore they choose  $s = 1$ .

**Part 2:** If  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$ , then (12) is strictly greater than (19), and thus the elites prefer a non-centralized state to forming a centralized state and having citizens form their grand coalition against it. Therefore, they opt for  $s = 0$ .

**Part 3:** In this case, because  $\Phi(N) > \Phi^*(N)$ , the citizens will always form their grand coalition, thus the elites also prefer to do the same, and thus again  $s = 1$ . Finally, note also that this part covers the remaining case where  $\Phi^*(N) < 1$ , since then by Assumption 1  $\Phi(N) > \Phi^*(N)$ . □

The most interesting part of Proposition 3 is the second one, establishing that when  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$ , the elites prefer a non-centralized state. The logic for this result goes via the escalation and political agenda effects discussed above and the externality that citizens from region 1 create on other citizens when they form a coalition with them. Because  $\Phi(N) < \Phi^*(N)$ , when the state is not centralized, citizens from region 1 prefer not to form a coalition with other citizens because of the escalation effect — when they form a coalition, they know that the elites will respond by forming their own coalition and thus escalate the conflict. Because  $\Phi(N) > \Phi_*(N)$ , this reluctance to form a citizen coalition is, from an ex ante point of view, highly beneficial for elites, because of the aforementioned externality and the political agenda effect: whenever citizens form this coalition, other citizens benefit and elites lose out. Each elite group, therefore, benefits from this externality with probability  $(N - 1)/N$ . In addition, because of the political agenda effect, if the state were centralized, they would have to endure higher levels of citizen investments in conflict capacity. In contrast, when the state is centralized, citizens from region 1 will form a coalition with other regions, obviating these benefits on elites. This can be thought of as a *reverse escalation effect*, reflecting the desire of elites to avoid the political agenda effect that arises when citizens form the grand coalition in response to state centralization — the centralization of the state this time escalates the conflict on the side of the citizens, who recognize that they are facing a national, unified elite and thus become more willing to form their own grand coalition, which then puts in motion the political agenda effect. In what follows, to make the discussion more specific, we do not use the term reverse escalation effect, and simply refer to the wish of the elite to avoid citizens forming a coalition as the political agenda effect.

Parts 1 and 3 are straightforward in this light: when  $\Phi(N) < \Phi_*(N)$ , the elites actually prefer citizens to form their grand coalition because they are sufficiently powerful against the citizens that the conflict being between  $N$  regions is advantageous for them.<sup>20</sup> Conversely, when  $\Phi(N) > \Phi^*(N)$ , the agenda effect is sufficiently powerful that citizens will always form a coalition with other regions, and consequently, the elites prefer  $s = 1$ .<sup>21</sup> Moreover, in this case, any transfers from the elites to citizens take the form of public goods.

Parts 1-3 together imply a nonmonotonic relationship between  $\Phi(N)$  and state centralization. Note a limitation of this result, however. We do not know whether  $\Phi_*(N)$  is greater than one or less than one. In the latter case, part 1 can never take place, and thus the relationship becomes monotonic. This nonmonotonic relationship also implies nonmonotone comparative statics, which are summarized in the next corollary. In this and the subsequent corollaries, the references to “more likely” or “less likely” stand for expanding or contracting the part of the parameter space where the result applies.

**Corollary 3.** *The likelihood of a centralized state and any public goods being provided is potentially nonmonotone in the productivity of public good provision,  $\mu$ , and in the heterogeneity of the preferences of different regions over public goods,  $\lambda$ .*

**Proof.** Suppose we are in the case in which  $1 < \Phi_*(N) < \Phi^*(N)$ . Then a higher  $\mu$  increases  $\Phi(N) = (N - (N - 1)\lambda)\mu$ , thus contracting the region of the space of the remaining parameters for which  $\Phi(N) < \Phi_*(N)$ . This makes state centralization and the provision of public goods less likely (the latter since when we are in part 2 of Proposition 3, any payments from the elite to citizens take the form of direct transfers, not the provision of public goods). Further increases in  $\Phi(N)$  ultimately take society out of the region where  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$ , thus again inducing state centralization and public good provision. The nonmonotonicity with respect to  $\lambda$  is similar. □

This corollary shows that, somewhat paradoxically, an increase in the productivity or value of public goods can make them less likely to be provided and the centralized state that can effectively provide them less likely to emerge.<sup>22</sup> This is because a higher  $\mu$  raises the marginal value of funds to citizens,  $\Phi(N)$ , which may then shift from the region in which  $\Phi(N) < \Phi_*(N)$  to the one where  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$ . This shift would induce a stronger political agenda effect, and the desire of elites to avoid this political agenda effect would push them towards not establishing a centralized state, which would in turn discourage citizens from forming a national coalition. Intuitively, a higher productivity of public goods makes the political agenda effect more powerful, and this further increases the strategic value of choosing a non-centralized state for the elites. Similarly, lower heterogeneity in the preference of different regions also has a nonmonotonic effect (whereas one might have expected that greater heterogeneity would always make public good provision less likely). The intuition is the same as for the comparative static for the productivity of public goods: as  $\lambda$

<sup>20</sup> In particular, in this region of the parameter space  $\theta^c$  is sufficiently large relative to  $\theta^*$  that  $H(N\theta^*(N, N, \Phi^*(N)) - N\theta^c)$  is much smaller than  $H(\theta^*(1, 1, 1) - \theta^c)$ , making the elite prefer the conflict to be between their grand coalition and the grand coalition of the citizens. Thus, as in e.g. Wittfogel (1957), it is the elites that gain from state centralization.

<sup>21</sup> Notice that here we are making use of the convention we introduced in Remark 2 that whenever the elites prefer to form their grand coalition, we will refer to this as  $s = 1$ , even though they could alternatively achieve this by choosing  $s = 0$ , and then ex post forming their grand coalition (with entirely equivalent consequences to  $s = 1$ ).

<sup>22</sup> If we adopted the generalization suggested in Remark 1, an increase in  $f(N)$  relative to  $f(1)$ , reflecting a greater (cost) advantage of the centralized state in providing public goods, we would also have nonmonotonic effects.

decreases, we may shift from the region in which  $\Phi(N) < \Phi_*(N)$  to the one where  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$ .

#### 4.2. Extension: Multiple equilibria

In this subsection, we relax Assumption 2 and replace it with the following:

**Assumption 2’.** The density of the distribution function  $H$ ,  $h$ , exists over its entire support  $\mathcal{H} \supset [-N\theta^e, 0]$ , is continuously differentiable and single-peaked.

There are two generalizations in Assumption 2’ relative to Assumption 2. First, it allows  $h$  to be single-peaked rather than nonincreasing everywhere. Second, the assumption that  $h(-N\theta^e)\alpha Y > \Gamma'(0)$  is dropped (we will analyze both this and the converse case in the next proposition).

Now the first-order condition of citizens in their conflict capacity, (4), may no longer have a unique solution. The next proposition delineates the circumstances under which this will be the case. When there is a unique equilibrium, the results are similar to those highlighted in Proposition 3. When there are multiple solutions, these translate into multiple equilibria, some of which have greater investments in conflict capacity by citizens. In this case, the patterns of state centralization are somewhat richer, but have the same overall intuition. To limit attention to the more interesting cases, whenever there are multiple equilibria, we focus on symmetric equilibria (when the equilibrium is unique, it is always symmetric as already characterized).

**Proposition 4.** Let  $\Phi^*(N)$  be as defined in Proposition 1, and suppose that Assumption 1 and 2’ hold.

1. Suppose also that  $h(-N\theta^e)\alpha Y > \Gamma'(0)$ ,  $h(\cdot)$  is concave to the left of its peak, and  $\Gamma'(\cdot)$  is weakly convex. Then, there exists a unique equilibrium. In this unique equilibrium, there exists  $\tilde{\Phi}(N)$  such that
  - (a) If  $\tilde{\Phi}(N) < \min\{\Phi_*(N), \Phi^*(N)\}$ , then the elites choose state centralization ( $s = 1$ ).
  - (b) If  $\Phi_*(N) < \Phi^*(N)$  and  $\Phi_*(N) < \tilde{\Phi}(N) < \Phi^*(N)$ , then the elites choose a non-centralized state ( $s = 0$ ).
  - (c) If  $\Phi^*(N) < \tilde{\Phi}(N)$ , then the elites choose state centralization ( $s = 1$ ).
2. Suppose instead that  $h(-N\theta^e)\alpha Y < \Gamma'(0)$ ,  $h(\cdot)$  is concave to the left of its peak and  $\Gamma'(\cdot)$  is weakly convex. Then, generically, there are either one or three symmetric equilibria. One equilibrium involves the elites choosing state centralization (i.e.,  $s = 1$ ) and no investment by the citizens in their conflict capacity, and is stable under best-response dynamics. In addition, when there are three equilibria, there is also a middle equilibrium, which is unstable under best-response dynamics, and an equilibrium with the highest level of citizen investment in their conflict capacity, which is also stable under best-response dynamics. Assuming that there are no equilibrium switches in response to changes in  $\Phi(N)$ , the equilibrium with the highest level of citizen investment is identical to the one described in part 1.
3. If  $h(\cdot)$  is not concave to the left of its peak or  $\Gamma'(\cdot)$  is not weakly convex, there is generically an odd number of equilibria. If  $h(-N\theta^e)\alpha Y < \Gamma'(0)$ , there always exists an equilibrium with state centralization and no investment by citizens in their conflict capacity. Assuming that there are no equilibrium switches in response to changes in  $\Phi(N)$ , the equilibrium with the highest level of citizen investment is identical to the one described in part 1.

**Proof.** First suppose that  $h(-N\theta^e)\alpha Y > \Gamma'(0)$  (which was imposed as part of Assumption 2 before) and that  $h(\cdot)$  is concave to the left of its peak and  $\Gamma'(\cdot)$  is weakly convex. Then, Figure 2a shows the marginal benefit and cost for additional investment for citizens that are part of the active coalition of citizens. The figure makes it clear that in this case there will be a single intersection, and thus the uniqueness result continues to hold. The rest of the proof is similar to that of Proposition 3, with the major difference that when  $h' > 0$ , we can no longer deduce that  $\Phi_*(N) < \Phi^*(N)$  when  $\Phi^*(N) \geq 1$  (but the case where  $\Phi_*(N) > \Phi^*(N)$  does not generate additional equilibrium outcomes, since when  $\tilde{\Phi}(N) > \Phi^*(N)$ , the citizens will form their grand coalition with or without a centralized state, and whether (12) is strictly greater than (19) becomes irrelevant).

Second, note that when  $h(-N\theta^e)\alpha Y < \Gamma'(0)$ ,  $h(\cdot)$  is concave to the left of its peak, and  $\Gamma'(\cdot)$  is weakly convex, Figure 2b shows that the marginal cost of investment may be everywhere above the marginal benefit for citizens from a given region, in which case there will be a unique equilibrium with no citizen investment. Figure 1c illustrates how, when this is not the case, there will exist three symmetric equilibria under state centralization, one of which is again the corner equilibrium with no citizen investment. In this corner equilibrium, because under state centralization ( $s = 1$ ), citizens choose no investment in their conflict capacity, and thus their probability of success is always lower under state centralization than any other coalition configuration (i.e.,  $h(-N\theta^e) < h(n\theta^e(n, n, \Phi(n)) - N\theta^e)$  for any  $n$ ), centralizing the state is a strict best response for the elites. The fact that the middle equilibrium is unstable under best-response dynamics follows from standard arguments (from the fact that marginal cost is less steep than marginal benefits). The equilibrium with the highest level of investment in citizen conflict capacity satisfies exactly the same first-order conditions for citizens as the unique equilibrium in part 1 under the usual assumption that there are no switches across equilibria in response to changes in parameters (compare Figures 2a and 2c).

Figure 2d shows how there might exist two equilibria in the non-generic case in which the  $\Gamma'$  schedule is tangent to  $h(\cdot)$ . Our notion of genericity rules out such tendencies (any small perturbation of either the  $\Gamma'$  or the  $h$  schedule in this case would restore the existence of one or three equilibria).

Finally, Figure 2e depicts a case with five equilibria when  $h(\cdot)$  is not concave to the left of its peak. The fact that the number of equilibria is generically odd follows from standard arguments. The equilibrium with the highest level of investment in their conflict capacity again satisfies the same first-order conditions for citizens.  $\square$

There are several senses in which this proposition is a generalization of Proposition 3. First, under the assumption that

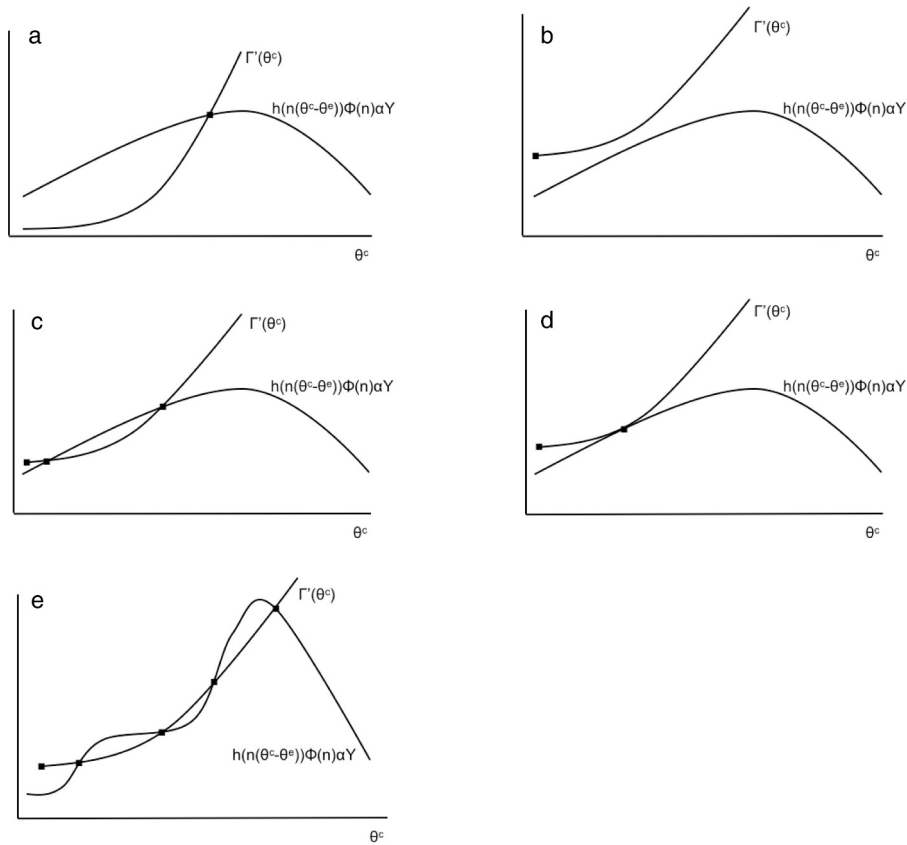


Fig. 2.

$h(-N\theta^c)\alpha Y > \Gamma'(0)$  (which was imposed via Assumption 2 previously), the result that there exists a unique equilibrium also holds when the density  $h$  is increasing provided that it is concave to the left of its peak and  $\Gamma'(\cdot)$  is weakly convex. However, in this case the threshold  $\Phi_*(N)$  may be greater than the threshold  $\Phi^*(N)$  even when  $\Phi^*(N) \geq 1$  (when the unique equilibrium is in the range where  $h' > 0$ ). When this is so, we also need to consider some additional configurations. In particular, we might have that  $\tilde{\Phi}(N) < \Phi^*(N) < \Phi_*(N)$ , but the implications of this are identical to the configuration where  $\tilde{\Phi}(N) < \Phi_*(N) < \Phi^*(N)$  (as summarized in part 1 of Proposition 3). Or we might have that  $\Phi^*(N) < \tilde{\Phi}(N) < \Phi_*(N)$ , and yet the proposition shows that in this case the implications are identical to the third part of Proposition 3, because citizens will form their grand coalition regardless (because  $\Phi^*(N) < \tilde{\Phi}(N)$ ).

Second and more radically, when  $h(-N\theta^c)\alpha Y < \Gamma'(0)$ , we are likely to have multiple equilibria, at least under state centralization.<sup>23</sup> In this case, as already noted, we focus on symmetric equilibria (where all regions that are part of the active coalition choose the same level of investment). The proposition shows that, unless the marginal cost of investment is everywhere above the marginal benefit, there will be three such equilibria when  $h(\cdot)$  is again concave to the left of its peak and  $\Gamma'(\cdot)$  is weakly convex. One of these equilibria will feature zero investment by the citizens under state centralization, and because this is the best configuration for the elites, they will have no incentive to deviate from state centralization. There will also exist a middle equilibrium which is unstable. Finally, the third symmetric equilibrium, the one with the highest level of investment by citizens in their conflict capacity, is similar to the equilibrium characterized in the first part of the proposition (and once again does not allow us to rank the thresholds  $\Phi^*(N)$  and  $\Phi_*(N)$ ).

Third, when  $h(\cdot)$  is not concave to the left of its peak or  $\Gamma'(\cdot)$  is not weakly convex, there can be more equilibria, though these have a similar structure to those characterized in part 2.

Overall, we interpret these results as showing the general robustness of the results highlighted in Proposition 3, but they also point out to additional insights. In particular, the multiplicity of equilibria might be one of the factors in play in accounting for the variation across societies in the patterns of state centralization and effectiveness of the state in social conflict.

<sup>23</sup> There may not be multiple equilibria without state centralization, since we may have that  $h(-\theta^c)\alpha Y > \Gamma'(0)$ , even though  $h(-N\theta^c)\alpha Y < \Gamma'(0)$ .

### 5. The role of a social democratic party

The main reason why there is no state centralization when  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$  is because the elites are using lack of state centralization as a strategic tool to discourage citizens from banding together in a national organization. They are able to do so because of two features of the setup studied here: first, they have the first mover advantage in choosing state centralization; and second, they recognize that once the identity of the region where citizens are strong is realized, those citizens (citizens from region 1) do not internalize the positive effect they have on the utility of citizens from other regions.

In this section, we argue that, under certain historical conditions, the emergence of a social democratic party might play a beneficial role countering this strategic incentive of elites. In such cases, social democratic politics will coordinate citizen conflict against elites, induce state centralization and pave the way to greater state capacity and the provision of general-interest public goods, and as such, improve the ex ante welfare of the citizens.

In Sweden, Norway and Denmark, social democratic parties emerged as relatively broad coalitions of different segments of society during the first half of the 20th century and played a critical role in coordinating and spearheading citizen demands from state institutions. The non-sectarian nature of Scandinavian social democrats is well illustrated by the Swedish case, where the party actively sought to include middle classes as well as its core supporters from the labor unions (see Berman, 2006, and also Acemoglu and Robinson, 2019). After 1932, it entered into what Gourevitch, (1986, p. 133) dubs the “cow trade” — a coalition with the agrarian party, which initially had very different objectives and priorities. Esping-Anderson (1985) emphasizes that such coalitions helped industrial workers become more cohesive in their negotiations with employers, while Berman (2006) points out that social democracy was built of communitarian impulses and tried to foment solidarity amongst workers. Moene and Wallerstein (2006) and Bengtsson (2019) emphasize that have suggested that the creation of social democracy in the 1930s, rather than following it, preceded many of the features of Scandinavian societies commonly argued to undergird social democratic politics, such as social harmony. Like our approach, this argument emphasizes how various societal and state institutions respond to the formation of a powerful social democratic party.

In the context of the model, we introduce a stage of the game preceding the state centralization decision, where citizens decide to form and delegate power to such a social democratic party. Because, as we have just argued, citizens may be better off ex ante from state centralization, there is a prima facie case that such an organization, by inducing centralization of bargaining (or of broader forms of conflict against the elites), might be useful for citizens. The next proposition shows that this is indeed the case, and in fact that a social democratic party would significantly expand the part of the parameter space where state centralization takes place and citizens would receive greater utility from the provision of public goods by elites.

**Proposition 5.** *Suppose that Assumptions 1 and 2 hold, and citizens have a decision to form a social democratic party before the state centralization decision of the elites. Then, there exists a unique equilibrium. Let  $\Phi^*(N) > 1$  and  $\Phi_*(N)$  be as defined in Proposition 3. There also exists  $\hat{\Phi}(N) < \Phi^*(N)$  such that:*

1. *If  $\Phi(N) \in (\hat{\Phi}(N), \Phi^*(N))$ , then citizens strictly prefer to form a social democratic party, and this induces state centralization on the part of the elites.*
2. *If  $\Phi(N) \in (\Phi_*(N), \hat{\Phi}(N))$ , then there is no state centralization.*
3. *If  $\Phi(N) < \Phi_*(N)$  or if  $\Phi(N) > \Phi^*(N)$ , then the elites choose  $s = 1$  regardless of the social democratic party formation decision of citizens.*

**Proof. Part 1.** The utilities of citizens without and with state centralization are, respectively,

$$U^{c*}[s = 0] = \frac{1}{N} [H(\theta^*(1, 1, 1) - \theta^e)\alpha Y - \Gamma(\theta^*(1, 1, 1))], \tag{22}$$

and

$$U^{c*}[s = 1] = H(N\theta^*(N, N, \Phi(N)) - N\theta^e)\Phi(N)\alpha Y - \Gamma(\theta^*(N, N, \Phi(N))). \tag{23}$$

When the citizens form a social democratic party, Lemma 2 implies that the elites will choose to centralize the state. Let us define the threshold  $\hat{\Phi}(N)$  such that (22) is equal to (23), i.e.,

$$\begin{aligned} & H(N\theta^*(N, N, \hat{\Phi}(N)) - N\theta^e)\hat{\Phi}(N)\alpha Y - \Gamma(\theta^*(N, N, \hat{\Phi}(N))) \\ &= \frac{1}{N} [H(\theta^*(1, 1, 1) - \theta^e)\alpha Y - \Gamma(\theta^*(1, 1, 1))]. \end{aligned}$$

Because the left-hand side of this expression is the same as the left-hand side of (11), while the right-hand side is equal to right-hand side of (11) divided by  $N$ , we have that  $\hat{\Phi}(N) < \Phi^*(N)$ . Moreover, because the left-hand side is monotonically increasing in  $\Phi(N)$ , we also have that whenever  $\Phi(N) > \hat{\Phi}(N)$ , citizens ex ante strictly prefer state centralization to a non-centralized state.

Suppose also that  $\hat{\Phi}(N) > \Phi_*(N)$ . Then, because  $\hat{\Phi}(N) \in (\Phi_*(N), \Phi^*(N))$ , in this range there will be no state centralization without a social democratic party (cfr. Proposition 3). Since citizens can induce state centralization by forming a social democratic party, they strictly prefer to do so in this range.

**Part 2.** In this range,  $\Phi(N) \in (\Phi_*(N), \hat{\Phi}(N))$ , citizens are better off with a non-centralized state, and so are the elites. Therefore, citizens choose not to form a social democratic party, and in response, the elite prefer not to centralize the state.

**Part 3.** This part directly follows from Proposition 3.  $\square$



Intuitively, the option to form a social democratic party turns citizens into Stackelberg leaders, and gives them an option to force the elites to centralize the state. As a result (and because  $\hat{\Phi}(N) < \Phi^*(N)$ ), state centralization becomes more likely with a social democratic party than without. Nevertheless, even in the presence of a social democratic party, the escalation effect is still present, and citizens may still shy away from inducing state centralization (in the proposition, this will be the case for the parameter values of  $\Phi(N)$  between  $\Phi_*(N)$  and  $\hat{\Phi}(N)$  if it happens that  $\Phi_*(N) < \hat{\Phi}(N)$ , which is a possibility since these two thresholds cannot be ranked unambiguously in general; in contrast if  $\Phi_*(N) > \hat{\Phi}(N)$ , citizens will always prefer to form a social democratic party and to induce state centralization).<sup>24</sup>

A straightforward corollary to Proposition 5 emphasizes that, compared to Proposition 3, state centralization is now more likely.

**Corollary 4.** *The likelihood of a centralized state and any public good being provided is more likely when the citizens have the possibility to form a social democratic party. The attractiveness of forming a social democratic party is increasing in the productivity of public good provision,  $\mu$ , and decreasing in the heterogeneity preferences of citizens in different regions,  $\lambda$ .*

**Proof.** The first part follows since  $\hat{\Phi}(N) < \Phi^*(N)$  and thus the parameter space where we have state centralization is larger. The second part follows as the gain in citizen utility with state centralization is increasing in  $\mu$  and decreasing in  $\lambda$ .  $\square$

This corollary thus highlights that the strategic incentive of the elites obtained in Corollary 3 was a consequence of the elites trying to strategically remove the political agenda effect. When the political agenda effect is already incorporated in the investment decisions of citizens by the presence of it social democratic party, this source of strategic incentive is weakened.<sup>25</sup>

### 6. Partial state centralization

In many societies such as Colombia, the Philippines or Pakistan, lack of full state centralization does not take the form of all elites and citizens engaging in local conflict, but certain subnational areas acting autonomously (both in terms of conflict and for public good provision decisions). For instance, in the Philippines the central government in Manila exercises very weak authority over several parts of the country, in particular, the southern Island of Mindanao where they have set up an autonomous region, currently called Bangsamoro (Abinales, 2000). The island has been plagued by insurgency since the 1960s (McKenna, 1998), and warlords have large private armies and operate with impunity (e.g. Human Rights Watch, 2010; Arguillas, 2011). The situation in Pakistan vis-à-vis the provinces of South and North Waziristan, Balochistan and the semi-autonomous tribal areas, which are all left to be ruled by local warlords and elites with minimal control of the national state, is similar.

In terms of our model, these situations correspond to partial state centralization which corresponds to a situation where there exists a well-defined national state, but this state does not extend its control to all the provinces, leaving areas such as Waziristan in Pakistan and Mindanao in the Philippines to largely autonomous local elites who themselves deal with local demands and conflicts. To capture this possibility in our model, suppose that in the first stage, instead of a simple decision over state centralization ( $s = 0$  or  $1$ ), the elites can form any partition of subcoalitions (e.g., if there are four regions, the first two and the last two could form to partially centralized local states). Citizens have access to the same technology of coalition formation as in our baseline model, whereby if region 1 (which is again designated as the strong region without loss of any generality) choose  $Z^c = 1$ , this will induce all other regions to join in. It is straightforward to see that Lemma 2 still holds, and thus the escalation effect continues to be present — citizens from region 1 know that if they choose  $Z^c = 1$ , then they will be facing the grand coalition of elites.

Our next result shows that in this case, they will often opt for partial state centralization, with detrimental results for citizens.

**Proposition 6.** *Suppose that Assumptions 1 and 2 hold. Let  $\Phi^*(N) > 1$  and  $\Phi_*(N)$  be as defined in Proposition 3, and recall that  $\Phi_*(N) < \Phi^*(N)$ . There now exists  $\Phi_2(N) \in (\Phi_*(N), \Phi^*(N))$ , such that if  $\Phi(N) \in (\Phi_*(N), \Phi_2(N))$ , then there will be partial state centralization and citizens will be worse off than in the case of no state centralization.*

**Proof.** When  $\Phi(N) < \Phi^*(N)$ , we have from (11) that

$$H(N\theta^*(N, N, \Phi(N)) - N\theta^c\Phi(N)\alpha Y - \Gamma(\theta^*(N, N, \Phi(N))) < H(\theta^*(1, 1, 1) - \theta^c)\alpha Y - \Gamma(\theta^*(1, 1, 1)).$$

Suppose instead that the elites form a subcoalition of two regions. This gives a new threshold  $\Phi_2(N)$  such that

$$H(N\theta^*(N, N, \Phi_2(N)) - N\theta^c\Phi_2(N)\alpha Y - \Gamma(\theta^*(N, N, \Phi_2(N))) = H(\theta^*(1, 2, 1) - 2\theta^c)\alpha Y - \Gamma(\theta^*(1, 2, 1)),$$

where  $\theta^*(1, 2, 1)$  refers to the investment level of citizens acting by themselves against a coalition of two regional elites (as captured

<sup>24</sup> Scott (2009) emphasizes how many communities in Southeast Asia adopted various strategies to stay outside of the control of the state, mostly to avoid exploitation.

<sup>25</sup> A question beyond the scope of our paper is why other political parties, such as communist or nationalist ones, have not played the same coordination role as social democratic parties. One possibility is that strong ideological priorities may make it difficult or impossible for a party to play this coordination role, which necessitates the formation of a broad coalition. Another possibility is that when such parties come to power, they or some of their leaders tend to attempt to take control of state institutions, undermining any role that they might have played in coordinating non-elite actions (see also Lipset et al., 1956).

by the fact that  $\theta^e$  is multiplied by 2). Notice that this is the level of investment by citizens when they face a coalition of two regional elites, but they have themselves not formed the grand coalition. By the same arguments as before, it is clear that  $\Phi_2(N) < \Phi^*(N)$ , because of the presence of 2 on the right-hand side of the above expression. Thus whenever  $\Phi(N) \in (\Phi_*(N), \Phi_2(N))$ , citizens will not form their national coalition when elites form a subcoalition of size two (and may not even form their national coalition in response to a subcoalition of a greater size, the threshold for which can be computed analogously).

Moreover, when this is the case, the elites will always strictly prefer to form such a subcoalition, because this reduces the probability that citizens win the conflict and thus the necessary transfers. To see this, recall that the threshold  $\Phi_*(N)$  is defined such that

$$H(N\theta^*(N, N, \Phi_*(N)) - N\theta^e) = \frac{1}{N}H(\theta^*(1, 1, 1) - \theta^e).$$

But we also have

$$\frac{1}{N}H(\theta^*(1, 1, 1) - \theta^e) > \frac{1}{N}H(\theta^*(1, 2, 1) - 2\theta^e). \tag{24}$$

To prove the latter, suppose to obtain a contradiction that this is not the case. Since  $h$  is nonincreasing from Assumption 2 this implies

$$h(\theta^*(1, 1, 1) - \theta^e) \geq h(\theta^*(1, 2, 1) - 2\theta^e).$$

In turn from (4) we have

$$\begin{aligned} h(\theta^*(1, 1, 1) - \theta^e)\alpha Y &= \Gamma'(\theta^*(1, 1, 1)), \text{ and} \\ h(\theta^*(1, 2, 1) - 2\theta^e)\alpha Y &= \Gamma'(\theta^*(1, 2, 1)). \end{aligned}$$

Given that  $\Gamma'' > 0$ ,  $\theta^*(1, 1, 1) \geq \theta^*(1, 2, 1)$ . This immediately implies that  $\theta^*(1, 1, 1) - \theta^e > \theta^*(1, 2, 1) - 2\theta^e$ , and thus that

$$H(\theta^*(1, 1, 1) - \theta^e) > H(\theta^*(1, 2, 1) - 2\theta^e),$$

yielding a contradiction. This implies that the two elites are strictly better off by partial state centralization (Note that their probability of facing strong citizens is given by  $2/N$ , and that if they face strong citizens they share the cost of concessions, and thus they strictly prefer a partial state when (24) holds).

Thus for  $\Phi(N) \in (\Phi_*(N), \Phi_2(N))$ , citizens do not form their grand coalition in the presence of partial state centralization of size 2 and the elites are better off in ex ante sense. As a result, partial state centralization of size 2 is preferred to no state centralization (which was the outcome without this option). This does not prove that the equilibrium will involve partial state centralization of size 2, since partial state centralization of some different size might be preferable for the elites. But it establishes that the equilibrium will have some partial state centralization as claimed. (And by the same arguments as above we can find the threshold  $\Phi_n(N) \in (\Phi_*(N), \Phi^*(N))$  where a partial state of  $n \in (1, N)$  elites forms which satisfies  $\Phi(N) < \Phi_n(N)$ , i.e. that the citizens do not form their grand coalition. In general the best response of the elites is to form the partial state with the highest  $n$  that satisfies this requirement).  $\square$

Intuitively, partial state formation takes place as a strategic step by elites to increase their power while still preventing the political agenda effect. Recall that when  $\Phi(N) \in (\Phi_*(N), \Phi^*(N))$ , the elites are able to, and prefer to, prevent the formation of a national coalition of citizens, thus obviating the political agenda effect. This, in particular, means that they cannot form a national state, because this will induce citizens to respond with their own national organization. However, when partial state formation is possible, the elites can take more limited steps to increase their organizational power and ability to fight demands from citizens, without inducing citizens to form a national organization. When they can do so, this not only leads to a pattern resembling the patchwork of subnational polities often acting outside of the control of the central state in many developing countries, but also further reduces the need to provide public goods to citizens to placate them.

### 7. Concluding remarks

The dominant view in political science is that many states remain incapable of centralizing power and monopolizing the means of coercion over the territory they ostensibly control, because they face insurmountable barriers to getting stronger. In this paper, we have argued that states may rather do so because they do not want to build that strength. At the center of our story is a new political economic force which we have called the *political agenda effect*. This effect captures the phenomenon that state centralization changes the nature of the societal conflict against the state and the political elites that control it. If this political agenda effect is sufficiently powerful, then the elites prefer to live with a non-centralized state (even if state centralization has other direct benefits, for example, the ability to confront local demands more effectively). The specific channel via which the nature of the conflict against the centralized state is different than against a non-centralized state in our model is the willingness of citizens with different interests to band together and shift their demands from parochial ones towards more general-interest ones (such as the provision of general-interest public goods). This increases the value of conflict to the citizens and encourages them to invest more in their conflict capacity, to the detriment of the elites who now have to placate stronger demands.

We developed these ideas in the context of the most parsimonious model we could construct which would still help us develop and elucidate the strategic forces at work. Several straightforward generalizations are possible but were not pursued in this paper to save space. We now briefly mention those before moving on to areas for future research which we view as more original and promising.

- Our analysis assumed that there were no direct benefits from a centralized state. This is clearly an unrealistic assumption. Centralized states bring a variety of benefits, ranging from more effective law enforcement and security to more efficient economy-wide regulation. As explained in Remark 1, introducing such benefits has no impact on our analysis or results, but would imply that strategic lack of state centralization will have greater social costs.
- Relatedly, an alternative natural assumption would have been that a centralized state also becomes more effective in providing public goods than a collection of local elites. In the context of our model, this would make the  $\Phi$  function also depend directly on whether the state is or is not centralized. It would also provide a direct mechanism via which a centralized state would be associated with public good provision. We chose not to make this assumption both on the grounds of parsimony and because the result that there will be an association between state centralization and effective public good provision is true in our model for a more interesting, endogenous reason — because without state centralization, specific, parochial transfers are the most economical way for the elite to meet citizen demands.
- An extension related to the previous point might be useful to develop. If a centralized state is necessary for public good provision (which is in contrast to our baseline model, where there is no technological benefit for public good provision from state centralization), the elites may strategically choose a non-centralized state even against the grand coalition of citizens, because this would be a strategic commitment to preventing public good provision, and when the resources captured in contest cannot be invested in public goods, the grand coalition of citizens becomes less effective. This extension is straightforward to develop and we did not do it to conserve space, even though we do find it interesting and potentially relevant for thinking about lack of state centralization in some contemporary cases.
- We also chose not to assume that a citizen organization can directly control the investment decisions of each of its constituent parts. But as shown in the Appendix, our (qualitative) results are identical when we introduce such control and thus remove the free-riding in citizen investment decisions.
- Yet another alternative assumption is to introduce economies of scale in the provision of public goods, which would go in the same direction of the regional spillovers we have used in the  $\Phi$  function. Once again, this would have no impact on our analysis or results.

We view our paper as part of a broader investigation. As already discussed in the Introduction, a few other papers have already proposed models in which states may strategically opt to remain weak (e.g., as a commitment not to expropriate, as in [Acemoglu, 2005](#); as a way of preventing rivals having an effective means of taxation, as in [Besley and Persson, 2009](#); or as an effort to prevent the formation of a powerful army capable of sharing rents, as in [Acemoglu et al., 2011b](#)). There are also several fruitful areas for future investigation. Here we mention a few.

- Our simplified modeling of heterogeneity across regions enabled a coalition of citizens, once formed, to formulate its demands effectively. Richer forms of heterogeneity would necessitate such organizations to mediate within-citizen conflicts and aggregate their heterogeneous preferences. Which types of coalitions or organizations can do so and how this interacts with the elite-citizen conflict and the state centralization decisions is a completely underresearched area.
- Also absent in our, and all other economic analyses to the best of our knowledge, is Habermas's notion, mentioned in the Introduction, that the formation of states paves the way for the emergence of 'public spheres,' which then impact the evolution of opinions. In a political economic setting, this can be interpreted as communication and information exchange improving following state centralization, which then impact how individuals would vote or make demands. Lack of state centralization can again emerge strategically in anticipation of different types of demands that will follow from this process.
- One of the most influential theories concerning state formation in social science is [Tilly's \(1995\)](#), linking state formation to war-making and the threat of war. The forces we have emphasized, as already anticipated in the Introduction, do not contradict this emphasis, but our analysis has abstracted from it. An interesting direction is to investigate the interplay between war and the strategic motives for opting for a non-centralized state. Beyond some obvious, but still interesting, comparative statics (e.g., showing that the forces we have emphasized become less important in the presence of the threat of war), such an extension could allow an investigation of whether some types of states with the capability to wage war while still strategically discouraging citizens from forming effective coalitions to make demands could emerge.
- Our analysis, like almost all other work in this topic in political economy, has abstracted from the internal organization of the state, which is an important and underresearched area. Centralized states behave differently than non-centralized ones in many dimensions (e.g., development of professional bureaucracies, adherence to the rule of law, etc.), which may be because the internal organization of a centralized state develops very differently than a non-centralized one. These considerations may also inform the extent to which state centralization is irreversible — after some amount of centralization, it may be impossible for the elite to reverse course. This is another important area that can be investigated in future work.
- Another interesting direction is to study why, as mentioned in footnote 25, it may have been difficult for other parties to play the same role of social democratic parties, and what other organizations and institutions also coordinate different citizen demands and thus supporting the process of state centralization.
- Last but not least, this area has very little empirical work (beyond those cited in the Introduction, which show the importance of state capacity). Whether centralized states change the nature of societal organization and how the anticipation of this impacts political equilibria are interesting, albeit difficult, areas to study empirically.

**Acknowledgment**

We thank editor Timur Kuran, two referees, seminar participants at the Canadian Institute for Advanced Research conference, CEMFI, Chicago-Northwestern Political Economy Conference, George Mason University, the Chr. Michelsen Institute, Nova School of Business and Economics Universidade De Lisboa, the University of East Anglia, the University of Oslo, and the University of Illinois, Urbana-Champaign for comments and suggestions, and Layne Kirshon for excellent research assistance. Acemoglu gratefully acknowledges financial support from the ARO MURI Award No. W911NF-12-1-0509.

**Appendix A. Omitted proofs and additional results**

In this Appendix we first present the proofs of Lemmas 1 and 2. We then turn to extensions of the model, where we first study the case of endogenous investments in the conflict capacity of elites, thereafter allow for subcoalition formation of citizens, and finally remove the free-rider effect in investment in conflict capacity.

*Omitted proofs*

**Proof of Lemma 1.** Suppose that the coalition of citizens is  $N^{z^c=1}$  with  $|N^{z^c=1}| = n$ , and is in conflict with a coalition of elites  $N^{z^e=1}$  with  $|N^{z^e=1}| = n^e$ . The citizens have at this stage invested in their conflict capacity, and denote their total conflict capacity by  $\tilde{\theta}^c$  and the total conflict capacity of the elites by  $\tilde{\theta}^e$  (in the case with exogenous conflict capacity of elites,  $\tilde{\theta}^e = n^e\theta^e$ , but  $\tilde{\theta}^e$  could be different than this when elites invest in their conflict capacity as we consider below, and this lemma applies in that case also).

Consider first the case where  $f_{N^{z^c=1}}^e = 1$ , i.e., where there is fighting. The cost of conflict capacity for each group of regional citizens is at this stage sunk and we simply denote it by  $\Gamma_i$  for group  $i$ , and thus the expected utility of each group of regional citizens  $i \in N^{z^c=1}$  from this conflict is given by

$$U_i^c \left[ f_{N^{z^c=1}}^e = 1 \right] = H(\tilde{\theta}^c - \tilde{\theta}^e)\Phi(n)\frac{n^e\alpha Y}{n} - \Gamma_i, \tag{A.1}$$

where  $H(\tilde{\theta}^c - \tilde{\theta}^e)$  is the probability the citizens win the fight,  $\Phi(n)$  the (maximum) marginal utility of funds for each group of regional citizens, and  $n^e\alpha Y/n$  the income gain for each group of regional citizens in the coalition if they win.

The expected utility of each elite  $i \in N^{z^e=1}$  from the conflict is similarly given by

$$U_i^e \left[ f_{N^{z^e=1}}^e = 1 \right] = [1 - H(\tilde{\theta}^c - \tilde{\theta}^e)]\alpha Y. \tag{A.2}$$

We next contrast this expression with the case where  $f_{N^{z^c=1}}^e = 0$ , i.e., where the elites make an offer that the citizen side prefers to fighting. Denote the utility that each citizen group  $i \in N^{z^c=1}$  obtains without fighting by  $U_i^c \left[ f_{N^{z^c=1}}^e = 0 \right]$ . For the citizens not to fight all members of the coalition must prefer to accept the offer from the elite coalition. This requires that the following participation constraint

$$U_i^c \left[ f_{N^{z^c=1}}^e = 0 \right] \geq U_i^c \left[ f_{N^{z^c=1}}^e = 1 \right]$$

should hold for each  $i \in N^{z^c=1}$ . Since utility is transferable across citizens of different regions (by choosing the level of transfers and public good provision), we can combine these participation constraints as

$$\sum_{i \in N^{z^c=1}} U_i^c \left[ f_{N^{z^c=1}}^e = 0 \right] \geq \sum_{i \in N^{z^c=1}} U_i^c \left[ f_{N^{z^c=1}}^e = 1 \right].$$

Since defeat has identical conclusions for all regional elites, as also noted in the text, all  $n^e$  members of the coalition of elites contribute equally, and denoting the total costs of concessions to the citizens by  $B$ , the symmetric maximization problem for each regional elite is

$$\max_{B \geq 0} Y - \frac{B}{n^e} \text{ subject to } \sum_{i \in N^{z^e=1}} U_i^e \left[ f_{N^{z^e=1}}^e = 0 \right] \geq \sum_{i \in N^{z^e=1}} U_i^e \left[ f_{N^{z^e=1}}^e = 1 \right],$$

where substituting for the elite coalition’s transfers, we also have

$$\sum_{i \in \mathcal{N}^{z^c=1}} U_i^c \left[ f_{\mathcal{N}^{z^c=1}}^e = 0 \right] = \Phi(n)B - \sum_{i \in \mathcal{N}^{z^c=1}} \Gamma_i.$$

Next substituting for  $U_i^c \left[ f_{\mathcal{N}^{z^c=1}}^e = 1 \right]$ , this maximization can be expressed as

$$\max_{B \geq 0} Y - \frac{B}{n^e} \text{ subject to } \Phi(n)B - \sum_{i \in \mathcal{N}^{z^c=1}} \Gamma_i \geq H(\tilde{\theta}^c - \tilde{\theta}^e)\Phi(n)n^e\alpha Y - \sum_{i \in \mathcal{N}^{z^c=1}} \Gamma_i.$$

Since the joint utility of the elite coalition is decreasing in  $B$ , it will choose  $B$  to make the joint participation constraint of the citizen coalition hold with equality. This gives

$$B = H(\tilde{\theta}^c - \tilde{\theta}^e)n^e\alpha Y,$$

and substituting this into the utility of regional elites without fighting, we have that for each elite  $i \in \mathcal{N}^{z^e=1}$

$$U_i^e \left[ f_{\mathcal{N}^{z^e=1}}^e = 0 \right] = (1 - \alpha H(\tilde{\theta}^c - \tilde{\theta}^e))Y > (1 - H(\tilde{\theta}^c - \tilde{\theta}^e))\alpha Y = U_i^e \left[ f_{\mathcal{N}^{z^e=1}}^e = 1 \right],$$

which establishes that each regional elite  $i \in \mathcal{N}^{z^e=1}$  is strictly better off with policy concessions than with fighting, and completes the proof of the lemma.  $\square$

**Proof of Lemma 2.** Suppose that citizens have formed a coalition  $\mathcal{N}^{z^c=1}$  with  $|\mathcal{N}^{z^c=1}| = n \geq 1$  (as usual including region 1, which is without loss of any generality the region with strong citizens).

Consider first the elites in regions  $i \notin \mathcal{N}^{z^c=1}$ . If these elites join the coalition  $\mathcal{N}^{z^c=1}$ , they participate in giving policy concessions (or in fighting), while if they do not participate in the coalition they face weak citizens that will not be able to make any demands. Since, in view of the fact that  $h(\cdot)$  is strictly positive over the relevant domain, policy concessions always have to be strictly positive, and elites in regions  $i \notin \mathcal{N}^{z^c=1}$  are strictly better off by not joining the coalition  $\mathcal{N}^{z^c=1}$ . This establishes that  $\mathcal{N}^{z^e=1} \subset \mathcal{N}^{z^c=1}$ .

Consider next the elites  $i \in \mathcal{N}^{z^c=1}$  (where citizens have joined the coalition  $\mathcal{N}^{z^c=1}$  with the strong citizens from region 1). Let us denote the investment in conflict capacity of citizens from region  $i$  in  $\mathcal{N}^{z^c=1}$  by  $\theta_j^c$ , so that the total conflict capacity of the coalition of citizens is  $\sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c$ . Suppose that the elite coalition  $\mathcal{N}^{z^e=1}$  forms in response and has  $n^l \leq n$  members. The utility of the regional elite  $i \in \mathcal{N}^{z^e=1}$  without state centralization can then be written

$$U_i^e [n^l | s = 0] = \left[ 1 - \alpha H \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - n^l \theta^e \right) \right] Y,$$

which incorporates from Lemma 1 that there will be policy concessions rather than fighting.

Now to obtain a contradiction suppose that  $\mathcal{N}^{z^e=1} \neq \mathcal{N}^{z^c=1}$ , which implies  $n^l < n$ . Consider a subcoalition of elites  $\tilde{\mathcal{N}}^{z^e=1} \subset \mathcal{N}^{z^e=1} \subset \mathcal{N}^{z^c=1}$  with size  $n^m \leq n^l$  (if on the other hand  $n^m > n^l$ , the same argument applies with the two coalitions  $\tilde{\mathcal{N}}^{z^e=1}$  and  $\mathcal{N}^{z^e=1}$  swapped around). The utility of an elite  $i$  in  $\tilde{\mathcal{N}}^{z^e=1}$  is

$$U_i^e [n^m | s = 0] = \left[ 1 - \alpha H \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - n^m \theta^e \right) \right] Y \leq U_i^e [n^l | s = 0],$$

where the last inequality follows as  $n^m \leq n^l$  and  $H(\cdot)$  is increasing. To complete the proof, now consider a regional elite  $j$  in  $\tilde{\mathcal{N}}^{z^e=1}$  deviating and switching to join  $\mathcal{N}^{z^e=1}$ . The size of this new coalition  $\mathcal{N}^{z^e=1} \cup \{j\}$  will be  $n^l + 1$ , thus giving each of its members,  $i \in \mathcal{N}^{z^e=1} \cup \{j\}$  a utility of

$$U_i^e [n^l + 1 | s = 0] = \left[ 1 - \alpha H \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - (n^l + 1)\theta^e \right) \right] Y > U_i^e [n^l | s = 0],$$

where again the last inequality follows because  $H(\cdot)$  is increasing. This establishes that any elite group that is not part of the largest coalition, in this instance  $\mathcal{N}^{z^e=1}$ , would be strictly better off by joining it. Since  $\mathcal{N}^{z^e=1} \subset \mathcal{N}^{z^c=1}$  from the first part of the proof, this establishes that  $\mathcal{N}^{z^e=1} = \mathcal{N}^{z^c=1}$ , completing the proof.  $\square$



Endogenous conflict capacity of elites

We now extend the model to include endogenous conflict capacity also of the elites. In particular, we assume that the conflict capacity  $\theta_i^e$  of each regional elite  $i$  results from investments with a strictly increasing, convex and continuously differentiable cost function  $\Gamma_e(\theta_i^e)$ , with  $\Gamma_e(0) = 0$ , and  $\Gamma_e'(\theta_i^e), \Gamma_e''(\theta_i^e) > 0$  for  $\theta_i^e > 0$ . The timing is the same as in the main model, with the only difference being that citizens and elites make their conflict capacity investments simultaneously.

As already noted in its proof, Lemma 1 directly generalizes to this case. To see that Lemma 2 also generalizes, one needs to modify its proof so that the switch is always to the elite coalition with the greatest total power (rather than simply with the largest number of members). Using these results, we can now carry out an analogous analysis to that in the text to show that the results presented in the text generalize with only minor modifications.

Equilibrium without a centralized state

Given Lemma 1, elites will invest in their conflict capacity to reduce the policy concessions they have to make (since these reduce the benefit for fighting for citizens). Since from Lemma 2,  $\mathcal{N}^{z^c=1} = \mathcal{N}^{z^e=1}$ , we set  $|\mathcal{N}^{z^c=1}| = |\mathcal{N}^{z^e=1}| = n$ , and write the maximization problem for each elite  $i \in \mathcal{N}^{z^e=1}$  as

$$\max_{\theta_i^e \geq 0} \left( 1 - \alpha H \left( n\theta^c - \sum_{j \in \mathcal{N}^{z^e=1}} \theta_j^e \right) \right) Y - \Gamma_e(\theta_i^e),$$

where we have already imposed the equilibrium result that citizens in coalition  $\mathcal{N}^{z^c=1}$  have chosen the same level of conflict capacity, which is denoted by  $\theta^c$ . The first-order condition for each elite  $i \in \mathcal{N}^{z^e=1}$  can be found as

$$h \left( n\theta^c - \sum_{j \in \mathcal{N}^{z^e=1}} \theta_j^e \right) \alpha Y - \Gamma_e'(\theta_i^e) = 0, \tag{A.3}$$

with the second-order condition

$$- h' \left( n\theta^c - \sum_{j \in \mathcal{N}^{z^e=1}} \theta_j^e \right) \alpha Y - \Gamma_e''(\theta_i^e) < 0. \tag{A.4}$$

We continue to impose Assumption 2, and in particular that  $h'(\cdot) \leq 0$ . For the second-order conditions of elites, this implies that we need to have  $|h'(\cdot)|$  not too large, which we assume in the remainder of the Appendix. This also assumes that the solution for the elite is also unique given  $\theta^c$ . Moreover, as already noted in the text, the investment of citizens is increasing in  $\theta^e$  (because  $h'(\cdot) \leq 0$ ), and with the same reasoning, the investment of elites is decreasing in  $\theta^c$ . This implies that under  $h'(\cdot) \leq 0$  and  $|h'(\cdot)|$  sufficiently small so that the second-order condition always holds, the equilibrium is unique. Let us denote the unique equilibrium level of conflict capacity of elites in the coalition  $\mathcal{N}^{z^e=1}$  by  $\theta^{e*}(n, n)$ , where the first argument conditions on the size of the coalition of citizens and the second argument designates conditioning on the size of the coalition of elites.

Note also that the maximization problem (imposing equilibrium behavior by elites) and the resulting investment levels by citizens are the same as in equation (4) in the text, with the only difference being that  $\theta^e$  is now replaced by  $\theta^{e*}(n, n)$ . As already anticipated, the second-order conditions of citizens are always satisfied. For future reference in this Appendix, we repeat this condition here as

$$h \left( \sum_{j \in \mathcal{N}^{z^c=1}} \theta_j^c - n\theta^{e*}(n, n) \right) \Phi(n) \alpha Y - \Gamma'(\theta_i^c) = 0, \tag{A.5}$$

and denote its solution by  $\theta^*(n, n, \Phi(n))$  as in the text.

Let us next turn to the coalition decision of citizens from region 1, which are once again designated the strong ones without loss of any generality. Suppose first that  $\mathcal{N}^{z^c=1} = \{1\}$ . Then noting that in this case the size of the coalition is 1 and from Assumption 1,  $\Phi(1) = 1$ , the utility of citizens from region 1 becomes

$$U_1^{c*}[1|s = 0] = \alpha H(\theta^*(1, 1, 1) - \theta^{e*}(1, 1))Y - \Gamma(\theta^*(1, 1, 1)). \tag{A.6}$$

The utility of elites from region 1 in this case is given by

$$U_1^{e*}[1|s = 0] = [1 - \alpha H(\theta^*(1, 1, 1) - \theta^{e*}(1, 1))]Y - \Gamma_e(\theta^{e*}(1, 1)). \tag{A.7}$$

Note again that the most efficient way of transferring resources for the elite is a direct transfer, so there will be no investment in public goods. That is,  $T_1 = \alpha H(\theta^*(1, 1, 1) - \theta^{e*}(1, 1))Y$  and  $G_1 = 0$ .

Suppose next that  $\mathcal{N}^{z^c=1} = \mathcal{N}$ . Then the utility of citizens becomes

$$U^{c*}[N|s = 0] = \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^{e*}(N, N))\Phi(N)Y - \Gamma(\theta^*(N, N, \Phi(N))), \tag{A.8}$$

and similarly the utility of elites is

$$U^{e*}[N|s = 0] = [1 - \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^{e*}(N, N))]Y - \Gamma_e(\theta^{e*}(N, N)). \tag{A.9}$$

Since from Assumption 1, we have  $\Phi(N) > 1$ , we can immediately conclude that  $T_i = 0$  and  $G_i = \mu\alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^{e*}(N, N))Y$  for all  $i \in N$ .

As in the text, citizens from region 1 will prefer to form the grand coalition of citizens when their utility from the grand coalition, (A.8), is strictly greater than their utility from standing alone, (A.6). The former, (A.8), is increasing in  $\Phi(N)$  with exactly the same argument as in footnote 16, while the latter, (A.6), is independent of  $\Phi(N)$ . Therefore, once again there exists a value of  $\Phi(N)$ ,  $\Phi^*(N)$ , such that

$$\begin{aligned} H(N\theta^*(N, N, \Phi^*(N)) - N\theta^{e*}(N, N))\Phi^*(N)\alpha Y - \Gamma(\theta^*(N, N, \Phi^*(N))) \\ = H(\theta^*(1, 1, 1) - \theta^{e*}(1, 1))\alpha Y - \Gamma(\theta^*(1, 1, 1)), \end{aligned} \tag{A.10}$$

and whenever  $\Phi(N) > \Phi^*(N)$ , citizens from region 1 (the strong citizens) will prefer the grand coalition, and whenever  $\Phi(N) < \Phi^*(N)$ , they will prefer to act by themselves.

The coalition formation decision of citizens, and the resulting utility levels for citizens, are thus identical to the ones summarized in Proposition 1, with the only difference being that  $\theta^e$  is now replaced by  $\theta^{e*}(1, 1)$  when  $\Phi(N) < \Phi^*(N)$ , and is replaced by  $\theta^{e*}(N, N)$  when  $\Phi(N) > \Phi^*(N)$ . The corollary that determines the expected utility of elites in different regions of the parameter space, which will be useful in our analysis in the rest of the Appendix becomes:

**Corollary 5.** When  $\Phi(N) < \Phi^*(N)$ , the expected utility of elites is

$$U^{e*}[s = 0] = \left[ 1 - \frac{\alpha}{N}H(\theta^*(1, 1, 1) - \theta^{e*}(1, 1)) \right] Y - \frac{1}{N}\Gamma_e(\theta^{e*}(1, 1)). \tag{A.11}$$

When  $\Phi(N) > \Phi^*(N)$ , the expected utility of elites is

$$U^{e*}[s = 1] = [1 - \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^{e*}(N, N))]Y - \Gamma_e(\theta^{e*}(N, N)). \tag{A.12}$$

*Equilibrium under a centralized state*

The coalition decision of citizens can be analyzed similarly to the one without state centralization, and it continues to be the case that the choice for citizens from region 1 is once again simply between  $N^{z^c=1} = \{1\}$  and  $N^{z^c=1} = N$ . Because when they choose  $N^{z^c=1} = \{1\}$ ,  $\Phi(1) = 1$ , their overall utility under state centralization will be

$$U_1^{c*}[1|s = 1] = \alpha H(\theta^*(1, N, 1) - N\theta^{e*}(1, N))Y - \Gamma(\theta^*(1, N, 1)),$$

while when they choose  $N^{z^c=1} = N$ , their overall utility under state centralization will be

$$U^{c*}[N|s = 1] = \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^{e*}(N, N))\Phi(N)Y - \Gamma(\theta^*(N, N, \Phi(N))).$$

The comparison of these two expressions once again boils down to

$$\begin{aligned} H(N\theta^*(N, N, \Phi(N)) - N\theta^{e*}(N, N))\Phi(N)\alpha Y - \Gamma(\theta^*(N, N, \Phi(N))) \\ > H(\theta^*(1, N, 1) - N\theta^{e*}(1, N))\alpha Y - \Gamma(\theta^*(1, N, 1)). \end{aligned}$$

As in the main text, this condition always holds. To see this, first note that  $\theta^{e*}$  now responds to the changes in the size of the citizen coalition. In particular, from (A.3) it follows that  $\theta^{e*}$  is decreasing in the size of the citizen coalition, and thus  $\theta^{e*}(1, N) > \theta^{e*}(N, N)$ . Then with the same argument in the text holding a fortiori including this additional effect, we have that

$$\begin{aligned} H(N\theta^*(N, N, \Phi(N)) - N\theta^{e*}(N, N))\alpha Y - \Gamma(\theta^*(N, N, \Phi(N))) \\ > H(\theta^*(1, N, 1) - N\theta^{e*}(1, N))\alpha Y - \Gamma(\theta^*(1, N, 1)), \end{aligned}$$

which implies the desired condition given that  $\Phi(N) > 1$  and  $\theta^{e*}(1, N) > \theta^{e*}(N, N)$ . Thus the results are essentially identical to those in Proposition 2, and the expected utility of elites under state centralization is given by

**Corollary 6.** When there is state centralization ( $s = 1$ ), the equilibrium utility of elites is given by

$$U^{e*}[s = 1] = [1 - \alpha H(N\theta^*(N, N, \Phi(N)) - N\theta^{e*}(N, N))]Y - \Gamma_e(\theta^{e*}(N, N)). \tag{A.13}$$

*Equilibrium state centralization*

We now characterize the state centralization decision of elites by combining the results from Corollaries 5 and 6. The reasoning is similar to the analysis in the main text: the elites prefer state centralization if and only if (A.13) exceeds (A.11). Note that (A.13) is decreasing in  $\Phi(N)$ , while (A.11) is independent of  $\Phi(N)$ . Thus the elites (strictly) prefer state centralization if and only if  $\Phi(N) < \Phi_*(N)$ , where  $\Phi_*(N)$  is defined by

$$\begin{aligned} \alpha H(N\theta^*(N, N, \Phi_*(N)) - N\theta^{e*}(N, N))Y + \Gamma_e(\theta^{e*}(N, N)) \\ = \frac{\alpha}{N}H(\theta^*(1, 1, 1) - \theta^{e*}(1, 1))Y + \frac{1}{N}\Gamma_e(\theta^{e*}(1, 1)). \end{aligned} \tag{A.14}$$

The only difference from the analysis in the main text is that, because of the presence of the cost function of the elites in (A.14), it is no longer necessarily the case that  $\Phi^*(N) > \Phi_*(N)$  when  $\Phi^*(N) > 1$ .

**Proposition 7.** Let  $\Phi^*(N)$  be as defined in (A.10) and  $\Phi_*(N)$  as defined in (A.14), and suppose that Assumptions 1 and 2 hold (and that  $|h'|$  is not so large, so that the second-order condition of elites, (A.4), holds). Then:

1. If  $\Phi_*(N) < \Phi^*(N)$  we have
  - (a) If  $\Phi(N) < \Phi_*(N)$ , then the elites choose state centralization ( $s = 1$ ).
  - (b) If  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$ , then the elites choose a non-centralized state ( $s = 0$ ).
  - (c) If  $\Phi^*(N) < \Phi(N)$ , then the elites choose state centralization ( $s = 1$ ).
2. If  $\Phi_*(N) > \Phi^*(N)$ , then the elites choose a non-centralized state ( $s = 0$ ).

**Proof. Part 1:** We now have three possibilities. In part a, where  $\Phi(N) < \Phi_*(N)$ , (A.11) is strictly less than (A.13), and thus the elites prefer a centralized state (with citizens forming their grand coalition) to a non-centralized state, and choose  $s = 1$ . In part b, where  $\Phi_*(N) < \Phi(N) < \Phi^*(N)$ , then (A.11) is strictly greater than (A.13), and thus the elites prefer a non-centralized state to forming a centralized state (because a centralized state would induce citizens to form their grand coalition), and thus choose  $s = 0$ . In part c, where  $\Phi(N) > \Phi^*(N)$ , the citizens will always form their grand coalition, and in response, the elites will also do so; thus  $s = 1$ .

**Part 2:** In this case, where  $\Phi_*(N) > \Phi^*(N)$ , either the citizens or the elites (or both groups) prefer a centralized state, and since both groups can induce this possibility by forming their grand coalition, we have  $s = 1$ .  $\square$

It is also straightforward to see that we have a version of Corollary 3, showing that comparative statics with respect to the productivity of public goods or the degree of preference heterogeneity across regions are non-monotonic. Overall, this analysis shows that when the conflict capacity of elites is endogenous, our main results generalize with minimal qualifications.

*Subcoalition formation of citizens*

In the text, when the strong citizens organized, all weak citizens could join their coalition (unless the strong citizens decided to stand alone). Because citizens from weak regions always prefer joining a coalition involving the strong citizens, such organizations always lead to the formation of the grand coalition of citizens, and the relevant choice for strong citizens was between standing alone and forming the grand coalition. We now extend the basic model and assume that strong citizens can invite specific regions to join the coalition, thus enabling them to choose the exact size of the coalition, denoted by  $n$  in this part of the Appendix. (Since all other regions are symmetric, which specific regions are included in the coalition is immaterial). Therefore, the choice of strong citizens is over coalitions of sizes  $n = 1, 2, \dots, N$ .

In this case, though the general qualitative forces remain the same, the analysis becomes more involved. To make progress, we now analyze the case in which  $H$  is uniform and  $\Gamma$  is quadratic, and show that, under some additional mild conditions, there will never be a coalition other than the singleton of the strong citizens or the grand coalition of all citizens, and thus the results from the text continue to apply unchanged. We then study the case in which  $H$  is uniform but  $\Gamma$  is an arbitrary convex function to show that even when citizens choose to form an “interior” subcoalition (i.e., a subcoalition greater than a singleton and smaller than their grand coalition), the elites prefer a non-centralized state under similar conditions to those in the text.

Note first that Lemmas 1 and 2 continue to hold. This, in particular, implies that whenever strong citizens choose a coalition of size  $n$ , they know that they will be facing an elite coalition of size  $n$ , which will then make them an offer to cease conflict in return for their expected return from the conflict.

Let us simplify the analysis here by treating  $n$  as a continuous variable, which will enable us to derive a first-order condition for an interior subcoalition decision of citizens.

As in the text, payoffs of citizens from region  $i \in \mathcal{N}^{z^c=1}$  with  $|\mathcal{N}^{z^c=1}| = |\mathcal{N}^{z^e=1}| = n$  are given from (5) as

$$U_i^{c*}[n|s = 0] = \alpha H(n\theta^*(n, n, \Phi(n)) - n\theta^e)\Phi(n)Y - \Gamma(\theta^*(n, n, \Phi(n))). \tag{A.15}$$

Let us next replace Assumption 2 with the following:

**Assumption 2''.**

- (i)  $H(\theta) = h_1\theta + h_0$ ,
- (ii)  $\Gamma(\theta) = \theta + \frac{1}{2k}\theta^2$ ,
- (iii)  $h_1\alpha Y > 1$ , and
- (iv)  $h_1 \geq \frac{1}{5k}$ .

Part (iii) of the assumption ensures that the investment level of citizens will be positive over the support of  $H$ , while the last part also requires the density of  $H$  not to be too low.

The first-order conditions of citizens from region 1, (4), for investment choice gives

$$\theta^*(n, n, \Phi(n)) = \Gamma'^{-1}(h_1\Phi(n)\alpha Y) = k(h_1\alpha\Phi(n)Y - 1) \equiv \theta^*(\Phi(n)), \tag{A.16}$$

where we have simplified notation by defining  $\theta^*$  as a function of  $\Phi(n)$  only, since in this case,  $n$  only matters for  $\theta^*$  via its effect on  $\Phi(n)$ . This allows us to rewrite (A.15) as

$$U_i^{c*}[n|s = 0] = \alpha H(n\theta^*(\Phi(n)) - n\theta^e)\Phi(n)Y - \Gamma(\theta^*(\Phi(n))).$$

Then treating  $n$  as a continuous variable, the derivative of this which must equal zero in a first-order condition for an interior

subcoalition decision is

$$\frac{dU_1^{c*}[n|s = 0]}{dn} = h_1\alpha\Phi(n)Y\left(\theta^*(\Phi(n)) - \theta^e + n\frac{\partial\theta^*(\Phi(N))}{\partial\Phi(N)}\frac{d\Phi(n)}{dn}\right) + \alpha H(n\theta^*(\Phi(n)) - n\theta^e)Y\frac{d\Phi(n)}{dn} - \Gamma'(\theta^*(\Phi(n)))\frac{\partial\theta^*(\Phi(N))}{\partial\Phi(N)}\frac{d\Phi(n)}{dn}$$

for some  $n \in (1, N)$ . Now substituting for  $\Gamma'(\theta^*(\Phi(n)))$  from (A.16), this expression can be simplified to

$$\frac{dU_1^{c*}[n|s = 0]}{dn} = h_1\alpha\Phi(n)Y(\theta^*(\Phi(n)) - \theta^e) + \alpha H(n\theta^*(\Phi(n)) - n\theta^e)Y\frac{d\Phi(n)}{dn} + h_1\alpha\Phi(n)Y(n - 1)\frac{\partial\theta^*(\Phi(N))}{\partial\Phi(N)}\frac{d\Phi(n)}{dn}. \tag{A.17}$$

Note that the last two terms on the right-hand side are nonnegative. Thus, if  $\theta^*(\Phi(n)) > \theta^e$  then the payoff of the strong citizens in region 1 is increasing in the number of groups of regional citizens  $n$  in the coalition, and we have a corner solution at  $n = N$ . This already reveals that for an interior solution to the subcoalition decision we need  $\theta^e$  to be sufficiently large.

Consider next the case where  $\theta^*(\Phi(n)) < \theta^e$ . In this case  $dU_1^{c*}[n|s = 0]/dn$  starts out negative since at  $n = 1$ ,  $d\Phi(n)/dn = 0$ , and in (A.17), we are left only with the first, negative term. We will next show that there exists no local maximum for  $U_1^{c*}[n|s = 0]$  with  $n \in (1, N)$ , so that the optimal choice for citizens from region 1 will be either  $n = 1$  or  $n = N$ . Recall first that there exists a unique  $n^*$  such that  $\Phi(n) > 1$  for  $n > n^*$  but  $\Phi(n^*) = 1$ . Thus clearly  $dU_1^{c*}[n|s = 0]/dn < 0$  for  $n < n^*$ , and the relevant parameter region for coalition formation is:  $n > n^*$ . Here  $d\Phi(n)/dn = (1 - \lambda)\mu$ , and also from (A.16) we have  $\partial\theta^*(\Phi(N))/\partial\Phi(N) = kh_1\alpha Y$ . Therefore, assuming an interior solution and substituting for  $H(n\theta^*(\Phi(n)) - n\theta^e)$  with  $h_1(nk(h_1\alpha\Phi(n)Y - 1) - n\theta^e)$ , we can rewrite (A.17) as

$$\begin{aligned} \frac{dU_1^{c*}[n|s = 0]}{dn} &= h_1\alpha\Phi(n)Y(k(h_1\alpha\Phi(n)Y - 1) - \theta^e) \\ &\quad + h_1\alpha Y(nk(h_1\alpha\Phi(n)Y - 1) - n\theta^e + h_0)(1 - \lambda)\mu \\ &\quad + h_1\alpha\Phi(n)Y(n - 1)kh_1\alpha Y(1 - \lambda)\mu, \end{aligned} \tag{A.18}$$

and also,

$$\begin{aligned} \frac{d^2U_1^{c*}[n|s = 0]}{dn^2} &= 2h_1\alpha Y(k(h_1\alpha\Phi(n)Y - 1) - \theta^e)(1 - \lambda)\mu \\ &\quad + (2n - 1)kh_1^2\alpha^2Y^2(1 - \lambda)^2\mu^2 \\ &\quad + 2\Phi(n)kh_1^2\alpha^2Y^2(1 - \lambda)\mu < 0 \end{aligned} \tag{A.19}$$

as the second-order condition.

Now for a solution to (A.18) to be an interior subcoalition choice, (A.18), implies that we must have

$$\begin{aligned} \theta^e &= \frac{1}{\Phi(n) + (1 - \lambda)\mu n} [\Phi(n)k(h_1\alpha\Phi(n)Y - 1) + \\ &\quad [nk(h_1\alpha\Phi(n) - 1) + h_0](1 - \lambda)\mu + (n - 1)\Phi(n)kh_1\alpha Y(1 - \lambda)\mu] \\ &\equiv \Theta(n), \end{aligned} \tag{A.20}$$

where the second line defines  $\Theta(n)$ .

For any solution to (A.18) to be an interior subcoalition choice, we need the second-order condition (A.19) to be satisfied. The following condition is then sufficient for this second-order condition to be violated whenever (A.18) holds and to ensure that there exists no local maximum:

$$2h_1\alpha Y(k(h_1\alpha\Phi(n)Y - 1) - \Theta(n))(1 - \lambda)\mu + (2n - 1)kh_1^2\alpha^2Y^2(1 - \lambda)^2\mu^2 + 2\Phi(n)kh_1^2\alpha^2Y^2(1 - \lambda)\mu > 0.$$

Substituting from (A.20) and rearranging, this can be written as

$$\begin{aligned} &\Phi(n)k(h_1\alpha\Phi(n)Y - 1) + [nk(h_1\alpha\Phi(n)Y - 1) + h_0](1 - \lambda)\mu + (n - 1)\Phi(n)kh_1\alpha Y(1 - \lambda)\mu \\ < &\Phi(n)k(h_1\alpha\Phi(n)Y - 1) + k(h_1\alpha\Phi(n)Y - 1)n(1 - \lambda)\mu \\ &+ \left(n - \frac{1}{2}\right)\Phi(n)kh_1\alpha Y(1 - \lambda)\mu + n\left(n - \frac{1}{2}\right)kh_1\alpha Y(1 - \lambda)^2\mu^2 \\ &+ \Phi(n)^2kh_1\alpha Y + \Phi(n)kh_1\alpha Yn(1 - \lambda)\mu. \end{aligned}$$

Now canceling the first term on the left-hand side with the first term on the right-hand side, the first part of the second term (excluding  $h_0$ ) with the second term on the right-hand side, and moving the third term on the left-hand side to the right-hand side and collecting terms with the third term and the last term on the right-hand side, we obtain the following sufficient condition:

$$h_0(1 - \lambda)\mu < \Phi(n)kh_1\alpha Y(1 - \lambda)\mu\left(n + \frac{1}{2}\right) + \Phi(n)^2kh_1\alpha Y + n\left(n - \frac{1}{2}\right)kh_1\alpha Y(1 - \lambda)^2\mu^2. \tag{A.21}$$

From the assumption that we have a symmetric uniform distribution,  $h_0 = \frac{1}{2h_1}$ , and from the last part of Assumption 2',  $\alpha Y h_1 > 1$ , and thus

$$\alpha Y > 2h_0.$$

Since  $\Phi(n) \geq 1$ ,  $n > 1$ , and  $(1 - \lambda)\mu < 1$ , the combination of the first two terms on the right-hand side is no less than  $\frac{5}{2}kh_1\alpha Y(1 - \lambda)\mu > 5kh_1h_0(1 - \lambda)\mu$ , and in view of the fourth part of Assumption 2', (A.21) is satisfied. This establishes:

**Proposition 8.** *Suppose Assumptions 1 and 2'' are satisfied. Then there will never be subcoalition formation of citizens, and the results from the text apply identically.*

Suppose next that the cost function is not quadratic, but satisfies the condition that  $\Gamma'(\theta) < \infty$  for all  $\theta \leq \theta^e + \varepsilon$  for some  $\varepsilon > 0$  (whereby this condition simply requires that citizens' investment in conflict capacity exceed  $\theta^e$ ), while  $H$  is still uniform. The analysis is similar in this case, except that we have

$$\theta^*(\Phi(n)) = \Gamma'^{-1}(h_1\Phi(n)\alpha Y),$$

and thus

$$\frac{d\theta^*(\Phi(n))}{d\Phi(n)} = \frac{h_1\alpha Y\Phi'(n)}{\Gamma''(h_1\Phi(n)\alpha Y)} = \frac{h_1\alpha Y(1 - \lambda)\mu}{\Gamma''(h_1\Phi(n)\alpha Y)}.$$

In fact,  $\theta^*(\Phi(n))$  is everywhere strictly increasing in  $\Phi(n)$ , and for  $\Phi(n)$  sufficiently large,  $\theta^*(\Phi(n)) > \theta^e$ .

Now rearranging the first-order condition of citizens from region 1 for subcoalition formation, (A.18), we obtain

$$\begin{aligned} \theta^e &= \frac{1}{\Phi(n) + n(1 - \lambda)\mu} [\Phi(n)\Gamma'^{-1}(h_1\Phi(n)\alpha Y) + n\Gamma'^{-1}(h_1\Phi(n)\alpha Y)(1 - \lambda)\mu \\ &\quad + (n - 1)\Phi(n)(1 - \lambda)\mu \frac{h_1\alpha Y(1 - \lambda)\mu}{\Gamma''(h_1\Phi(n)\alpha Y)} + h_0(1 - \lambda)\mu] \\ &\equiv \Theta(n). \end{aligned}$$

We can no longer rule out an interior subcoalition, but we can still show that lack of state centralization can emerge for a wide range of parameter values in this case. To do this, note that for an interior solution to exist, we need

$$\theta^e \geq \Theta(n)$$

for some  $n \in (1, N)$  (since otherwise (A.18) cannot be satisfied for any interior subcoalition, and we are back to the conclusion of Proposition 8). Next, it is straightforward to verify that  $\Theta(n)$  is increasing for  $n$  sufficiently large. This implies that there exists a maximal  $\bar{n}$  such that any subcoalition will be less than  $\bar{n}$  (where  $\bar{n}$  is defined such that  $\theta^e < \Theta(n)$  for all  $n > \bar{n}$ ). Note that this  $\bar{n}$  is independent of  $N$ .

Suppose this condition does not hold, so that there is a choice of interior coalition at  $\bar{n} \in (1, \bar{n}]$  (where the fact that the size of the subcoalition is less than  $\bar{n}$  follows from the previous paragraph). Then the question is whether the elites will still choose a non-centralized state. The condition for that is now

$$(1 - \frac{\bar{n}}{N}\alpha H(\bar{n}\theta^*(\Phi(\bar{n})) - \bar{n}\theta^e)Y > (1 - \alpha H(N\theta^*(\Phi(N)) - N\theta^e))Y, \tag{A.22}$$

for some choice of intermediate subcoalition  $\bar{n}$ . Rearranging this condition, we obtain

$$(N^2 - \bar{n}^2)\theta^e < h_1\alpha Y [N^2\Gamma'^{-1}(h_1\Phi(N)\alpha Y) - \bar{n}^2\Gamma'^{-1}(h_1\Phi(\bar{n})\alpha Y)].$$

But since  $\bar{n}$  is less than  $\bar{n}$ , which is independent of  $N$ , and because  $\theta^*(\Phi(N)) = \Gamma'^{-1}(h_1\Phi(N)\alpha Y)$  will exceed  $\theta^e$  for  $\Phi(N)$  sufficiently large, we have that this condition will necessarily be violated for  $N$  sufficiently large. Thus we have proved:

**Proposition 9.** *Suppose Assumption 1 is satisfied, and let  $H(\cdot)$  be uniform and suppose that there exists  $\varepsilon > 0$  such that  $\Gamma'(\theta) < \infty$  for all  $\theta \leq \theta^e + \varepsilon$ . Then, there exists  $\bar{N}$  such that for all  $N > \bar{N}$ , the elites will choose a non-centralized state ( $s = 0$ ) even if citizens form a subcoalition of size less than  $N$ .*

In contrast to Proposition 8, which generalizes the results from the text by establishing that citizens will not form an intermediate coalition, this result shows that, when the size of society (in terms of the number of regions or different groups) is sufficiently large, a non-centralized state will emerge as an equilibrium phenomenon even when citizens prefer to form an intermediate coalition.

*Removing the free-rider effect*

We now extend our baseline model by allowing citizen coalitions to jointly decide their investments in conflict capacity, which will remove the free-rider effect in investment decisions of citizens. Hence, the equivalent of (3) in the text now becomes

$$\max_{\{\theta_i^c\}_{i \in N} \sum_{i=1}^c=1} H \left( \sum_{j \in N} \sum_{z^c=1} \theta_j^c - n\theta^e \right) \Phi(n)\alpha nY - \sum_{i \in N} \sum_{z^c=1} \Gamma(\theta_i^c),$$



where recall that  $|\mathcal{N}^{z^c=1}| = n$ .

Notice two differences relative to (3). First, the value of victory in this conflict is now multiplied by  $n$ , which is the size of the coalition of citizens,  $N^{z^c=1}$ . This reflects the fact that the maximization is now from the viewpoint of the coalition, thus the return to winning the conflict is the total resources that the coalition will control following such a victory,  $nY$ . Second, the coalition also directly controls the decisions of all regions that are part of the coalition, as reflected by the fact that the maximization is over  $\{\theta_i\}_{i \in N^{z^c=1}}$ , and the total cost of investment of all of these groups is subtracted from the gross return. These two differences together imply that there is no longer the free-rider effect in the investment decisions (which was present in the text because each group of citizens in the coalition failed to internalize the benefit that it would create on other members of the coalition). As a consequence, the level of investment in conflict capacity by citizens in this case is always greater than the one in the text. More formally, denoting the level of investment in this case by  $\bar{\theta}^*(n, n, \Phi(n))$ , we have the equivalent of (4) in the text — the first-order condition that this investment satisfies — as

$$h(n\bar{\theta}^*(n, n, \Phi(n)) - n\theta^c)\Phi(n)\alpha Yn - \Gamma'(\bar{\theta}^*(n, n, \Phi(n))) = 0. \tag{A.23}$$

The comparison of this equation to (4) immediately yields

$$\bar{\theta}^*(n, n, \Phi(n)) > \theta^*(n, n, \Phi(n)).$$

The rest of the analysis proceeds as before, and the relevant choice for citizens from region 1 is once again to engage in conflict by themselves vs. to form the grand coalition. Which one they will prefer will again depend on whether  $\Phi(N)$  is above a critical threshold, which is now denoted by  $\bar{\Phi}^*(N)$ . Moreover, this threshold is again given by a condition that is essentially identical to (11), except that  $\bar{\theta}^*$  replaces  $\theta^*$ ; this captures the fact that the investment decisions of different coalitions will now satisfy the first-order condition (A.23) rather than (4). Namely:

$$\begin{aligned} H(N\bar{\theta}^*(N, N, \bar{\Phi}^*(N)) - N\theta^c)\bar{\Phi}^*(N)\alpha Y - \Gamma(\bar{\theta}^*(N, N, \bar{\Phi}^*(N))) \\ = H(\theta^*(1, 1, 1) - \theta^c)\alpha Y - \Gamma(\theta^*(1, 1, 1)) \end{aligned}$$

(where we have also incorporated the fact that  $\bar{\theta}^*(1, 1, 1) = \theta^*(1, 1, 1)$ , since there was no free-rider problem when citizens from region 1 were acting by themselves).

Since the monotonicity arguments are identical to those in the text, this condition then immediately shows that citizens from region 1 will choose to form their grand coalition whenever  $\Phi(N) > \bar{\Phi}^*(N)$ . Since  $\bar{\theta}^*(N, N, \Phi(N)) > \theta^*(N, N, \Phi(N))$ , we also have  $\bar{\Phi}^*(N) < \Phi^*(N)$ , which simply reflects the intuitive idea that when the free-rider problems are not present, citizens will invest more in a coalition, and thus forming the grand coalition becomes more attractive. A similar argument can also be used to define the analog of threshold  $\Phi_*(N)$ ,  $\bar{\Phi}_*(N)$ . Because the grand coalition of citizens is now investing more in their conflict capacity, state centralization becomes less attractive to elites. As a consequence, this new threshold satisfies  $\bar{\Phi}_*(N) < \Phi_*(N)$ . Using these thresholds and following identical steps to the proof of Proposition 3, we establish:

**Proposition 10.** *Suppose that citizen coalitions are not subject to the free-rider problem in their investments in conflict capacity. Then Proposition 3 holds with  $\Phi_*(N)$  replaced by  $\bar{\Phi}_*(N)$  and  $\Phi^*(N)$  replaced by  $\bar{\Phi}^*(N)$ .*

This proposition thus shows that, qualitatively, all of the results from the text apply in this case as well. Corollary 3 also applies identically.

## References

Abinales, P.N., 2000. Making Mindanao: Cotabato and Davao in the Formation of the Philippine Nation-State. Ateneo De Manila Univ. Press, Manila.

Acemoglu, D., 2005. Politics and economics in weak and strong states. *Journal of Monetary Economics* 52, 1199–1226.

Acemoglu, D., Garcia-Jimeno, C., Robinson, James A., 2015. State capacity and economic development: A network approach. *American Economic Review* 105 (8), 2364–2409.

Acemoglu, D., Moscona, J., Robinson, J.A., 2016. State capacity and american technology. *American Economic Review* 106 (5), 61–67.

Acemoglu, D., Robinson, J.A., 2019. *The Narrow Corridor: States, Societies, and the Fate of Liberty*. Penguin, New York.

Acemoglu, D., Ticchi, D., Vindigni, A., 2010. Persistence of civil wars. *Journal of the European Economic Association* 8 (2-3), 664–676.

Acemoglu, D., Ticchi, D., Vindigni, A., 2011. Emergence and persistence of inefficient states. *Journal of the European Economic Association* 9 (2), 177–208.

Amsden, A.H., 1989. *Asia's Next Giant: South Korea and Late Industrialization*. Oxford Univ. Press, New York.

Andaya, B.W., Andaya, L.Y., 1982. *A History of Malaysia*. Macmillan, London.

Anderson, B.R.O'G., 2001. *Violence and the State in Suharto's Indonesia*. Cornell Univ. Press, Ithaca.

Anderson, P., 1974. *Lineages of the Absolutist State*. Verso, London.

Anderson, S., François, P., Kotwal, A., 2015. Clientelism in indian villages. *American Economic Review* 105 (6), 1780–1816.

Arguillas, C.O., 2011. Maguindanao: The long shadow of the Ampatuans. In: Chua, Y.T., Rimban, L..D. (Eds.), *Democracy at Gunpoint*. The Asia Foundation, Makati City.

Bates, R.H., 1981. *Markets and States in Tropical Africa*. Univ. of California Press, Berkeley.

Bengtsson, E., 2019. The swedish sonderweg question: democratization and inequality in comparative perspective, c. 750-1920. *Past and Present* 244, 123–161.

Berman, S., 2006. *The Primacy of Politics: Social Democracy and the Making of Europe's Twentieth Century*. Cambridge Univ. Press, New York.

Besley, T., Coate, S., 2003. Centralized versus decentralized provision of public goods: A political economy approach. *Journal of Public Economics* 87, 2611–2637.

Besley, T., Persson, T., 2009. The origins of state capacity: Property rights, taxation and politics. *American Economic Review* 99 (4), 1218–1244.

Besley, T., Persson, T., 2011. *Pillars of Prosperity*. Princeton Univ. Press, Princeton.

Bolton, P., Roland, G., 1997. The break-up of nations: A political economy analysis. *Quarterly Journal of Economics* 112 (4), 1057–1090.

Brewer, J., 1990. *The Sinews of Power: War, Money and the English State, 1688-1783*. Harvard Univ. Press, Cambridge.

Crémer, J., Palfrey, T.R., 1999. Political confederation. *American Political Science Review* 93 (1), 69–83.

- Dincecco, M., Katz, G., 2016. State capacity and long-run economic performance. *Economic Journal* 126, 590, 189–218.
- Emerson, R., 1937. *Malaysia: A Study in Direct and Indirect Rule*. The Macmillan Company, New York.
- Epstein, S.R., 2000. *Freedom and Growth*. Routledge, London.
- Esping-Anderson, G., 1985. *Politics Against Markets: The Social Democratic Road to Power*. Princeton Univ. Press, Princeton.
- Evans, P.B., 1995. *Embedded Autonomy*. Princeton Univ. Press, Princeton.
- Evans, P.B., Rauch, J.E., 1999. Bureaucracy and growth: A cross-national analysis of the effects of 'Weberian' state structures on economic growth. *American Sociological Review* 64, 748–765.
- Evans, P.B., Rauch, J.E., 2000. Bureaucratic structure and bureaucratic performance in less developed countries. *Journal of Public Economics* 75, 49–71.
- de Figueiredo, Rui, Weingast, B.R., 2005. Self-enforcing federalism. *Journal of Law, Economics and Organization* 21, 103–135.
- Filippov, M., Ordeshook, P.C., Shvetsova, O., 2004. *Designing Federalism: A Theory of Self-Sustainable Federal Institutions*. Cambridge Univ. Press, New York.
- Gennaoli, N., Rainer, I., 2007. The modern impact of pre-colonial centralization in Africa. *Journal of Economic Growth* 12 (3), 185–234.
- Gourevitch, P.A., 1986. *Politics in Hard Times: Comparative Responses to International Economic Crises*. Princeton Univ. Press, Princeton.
- Habermas, J., 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. MIT Press, Cambridge.
- Hardling, P., 1996. The Waning of 'Old Corruption': The Politics of Economical Reform in Britain, 1779-1846. Oxford Univ. Press, New York.
- Harper, T.H., 1999. *The End of Empire and the Making of Malaya*. Cambridge Univ. Press, New York.
- Hechter, M., Brustein, W., 1980. Regional modes of production and patterns of state formation in western Europe. *American Journal of Sociology* 85 (5), 1061–1094.
- Hooper, V.M., 2003. *A Short History of Malaysia*. Allen & Unwin, London.
- Iliffe, J., 1979. *A Modern History of Tanganyika*. Cambridge Univ. Press, New York.
- Jean-Laurent, R., 1990. The development of irrigation in provence 1700-1860: The french revolution and economic growth. *Journal of Economic History* 50, 615–638.
- Johnson, C.A., 1982. *MITI and the Japanese Miracle*. Stanford Univ. Press, Stanford.
- Johnson, N. D., 2015. *Taxes, national identity, and nation building: Evidence from france*. [Http://noeldjohnson.net/noeldjohnson.net/Home.html](http://noeldjohnson.net/noeldjohnson.net/Home.html).
- Katznelson, I., 1985. Working class formation and the state: Nineteenth century england in american perspective. In: Evans, P.B., Dietrich, R., Theda, S. (Eds.), *Bringing the State Back*. Cambridge Univ. Press, New York.
- Kitschelt, H., 2000. Linkages between citizens and politicians in democratic polities. *Comparative Political Studies* 33 (6/7), 845–879.
- Kitschelt, H., Wilkinson, S.I., 2007. *Patrons, Clients and Policies: Patterns of Democratic Accountability and Political Competition*. Cambridge Univ. Press, New York.
- Lipset, S., Martin, T., Martin, C., Samuel, J., 1956. *Union Democracy: The Internal Politics of the International Typographical Union*. Free Press, New York.
- Lizzeri, A., Persico, N., 2001. The provision of public goods under alternative electoral incentives. *American Economic Review* 91 (1), 225–239.
- Lizzeri, A., Persico, N., 2004. Why did the elites extend the suffrage? democracy and the scope of government, with an application to britain's 'age of reform. *Quarterly Journal of Economics* 119 (2), 707–765.
- Lockwood, B., 2002. Distributive politics and the costs of centralization. *Review of Economic Studies* 69 (2), 313–337.
- McAdam, D., McCarthy, J.D., Zald, M.N., 1988. Social movements. In: Smelser, N.J. (Ed.), *The Handbook of Sociology*. Sage Beverly Hills.
- McKenna, T.M., 1998. Muslim rulers and rebels: Everyday politics and armed separatism in the Southern Philippines. Univ. of California Press, Berkeley.
- Michalopoulos, S., Papaioannou, E., 2013. Pre-colonial ethnic institutions and contemporary african development. *Econometrica* 81 (1), 113–152.
- Migdal, J., 1988. *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*. Princeton Univ. Press, Princeton.
- Migdal, J., 2001. *State-in-Society: Studying How States and Societies Transform and Constitute One Another*. Cambridge Univ. Press, New York.
- Moene, K., Michael, W., 2006. Social democracy as a development strategy. In: Bardhan, P., Bowles, S., Wallerstein, M. (Eds.), *Globalisation and Egalitarian Redistribution*. Russel Sage Foundation, New York.
- North, D.C., 1981. *Structure and Change in Economic History*. W.W. Norton & Co, New York.
- Nye, J., 2007. *War, Wine and Taxes: The Political Economy of Anglo-French Trade, 1689–1900*. Princeton Univ. Press, Princeton.
- O'Neill, K., 2005. *Decentralizing the State: Elections, Parties and Local Power in the Andes*. Cambridge Univ. Press, New York.
- Osafo-Kwaako, P., Robinson, J.A., 2013. Political centralization in pre-colonial Africa. *Journal of Comparative Economics* 41 (1), 534–564.
- Riker, W.H., 1964. *Federalism: Origin, Operation, Significance*. Little, Brown and Company, Boston.
- Robinson, J.A., Verdier, T., 2013. The political economy of clientelism. *Scandinavian Journal of Economics* 115, 260–291.
- Saylor, R., 2014. *State Building in Boom Times: Commodities and Coalitions in Latin America and Africa*. Oxford Univ. Press, New York.
- Scott, J.C., 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale Univ. Press, New Haven.
- Scott, J.C., 2009. *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. Yale Univ. Press, New Haven.
- Shefter, M.A., 1977. Party and patronage: Germany, england and italy. *Politics & Society* 7 (4), 403–451.
- Shefter, M.A., 1993. *Political Parties and the State: The American Historical Experience*. Princeton Univ. Press, Princeton.
- Slater, D., 2010. *Ordering Power: Contentious Politics and Authoritarian Leviathans in Southeast Asia*. Cambridge Univ. Press, New York.
- Tilly, C., 1978. *From Mobilization to Revolution*. Addison-Wesley, New York.
- Tilly, C., 1990. *Coercion, Capital and European States, AD 990–1990*. Basil Blackwell, Cambridge.
- Tilly, C., 1995. *Popular Contention in Great Britain, 1758-1834*. Harvard Univ. Press, Cambridge.
- Human Rights Watch, 2010. *They own the people: The ampatuans, state-backed militias, and killings in the southern philippines*. <https://www.hrw.org/report/2010/11/16/they-own-people/ampatuans-state-backed-militias-and-killings-southern-philippines>.
- Wittfogel, K.A., 1957. *Oriental Despotism: A Comparative Study of Total Power*. Yale Univ. Press, New Haven.