The Construction of Political Order

Scott F. Abramson* Emiel Awad† Brenton Kenkel‡

January 15, 2019

Abstract

We develop a model to understand the kinds of peaceful political order that can emerge from anarchy. We characterize existence of three types of peaceful order: 1) *peaceful states of nature* where no agent invests in coercive force; 2) *monopolies of violence* where a single agent invests in the production of force; and 3) *balances of power* where all agents invest in coercion. We show that the welfare-maximizing peaceful state of nature is sustainable only if conflict would destroy all of society's wealth. Additionally, we find that it is more costly to sustain political order through a monopoly of force than to have multiple agents maintain coercive abilities. Lastly, we show that the political order most preferred by any individual agent entails an unnecessarily high investment in coercion, larger than is strictly required to maintain peace.

Word count: 9,927

*Assistant Professor, Department of Political Science, University of Rochester.

†Ph.D. candidate, Department of Political Science, University of Rochester.

‡Assistant Professor, Department of Political Science, Vanderbilt University.

1 Introduction

In the contemporary world, complex hierarchical institutions that control the use of force—states—are the ubiquitous form of political organization. This is a relatively recent phenomenon. Twelve thousand years ago, humans inhabited the entire globe (besides Antarctica), yet nowhere were there political structures more complex than small bands or clan-based groups. In 4000 BC the first entities recognized as states emerged in Mesopotamia, and by 1500 BC states came to dominate the Nile, Indus, and Yellow River valleys. Still, by the time of the Columbian Exchange, all of Australia, large parts of the Americas and Africa, and smaller portions of Europe and Asia were devoid of state institutions.[1]

It is common to view the transition away from orderless, amorphic, society as a boon to human welfare. Indeed, there is broad consensus among development professionals and academics that statelessness is anathema to growth, security, and the protection of human rights (Rotberg 2002; Levy and Kpundeh 2004; Bates 2008; Besley and Persson 2011). Yet the archaeological record belies this conventional wisdom. By most measures of well-being, the earliest states were harmful for welfare (Cohen 1989; Larsen 1995; Edgerton 2010). Anthropological scholarship suggests that these harms are driven by negative ecological and biological externalities associated with the formation of early states. Maladies generated by sedentism, increased population density, and lower variability in diets are argued to have made life far more nasty, brutish, and short for those living in early states than for those living outside of them.[2]

In this paper we develop a model to understand the kinds of peaceful political orders that are supportable. We show that whenever there is a peaceful equilibrium with a monopoly of violence, there is also a social welfare–enhancing peaceful equilibrium in which multiple agents maintain coercive capabilities. In contrast with the prior anthropological literature,

[1]On the timing described above, see Sandeford (2018).

[2]For a synthesis of the anthropological literature on the topic, see Scott (2017).

we generate this result not by asserting that states provide a public bad. We suggest that political economy rather than epidemiology may be responsible for this reduction in well-being. Lower levels of social welfare result from the high levels of unproductive investment in coercive abilities required for a single agent to deter conflict over economic output. Moreover, we show that a monopolist of violence will invest more in coercive capability than is strictly necessary to maintain order, as doing so allows the monopolist to extract more despite reducing society's overall wealth.

The players in our game begin in anarchy, where there is no third party to enforce property rights and each player can use force to appropriate others' wealth. We examine the ability of these actors to construct institutions as "formal rules of the game" (North 1990, p. 3) with two characteristics. First, we want to know when agents in an institution-free society can develop rules that *prevent the use of violence*. Second, we seek to understand when these rules are *self-enforcing*. In other words, in an environment where agents can resort to violence and can potentially flee the imposition of political order, we want to know when it is in the individual interests of each agent to participate in the institution and refrain from violence.

We characterize the conditions necessary to sustain an institution-free *peaceful state of nature* wherein peace prevails even though no one invests in coercive force. When peace of this sort is unsustainable, we describe the conditions that allow for self-enforcing political institutions that prevent all violence but whose participants nevertheless make costly investments in coercive abilities. There are two types of order underpinned by coercion. First, there is a *monopoly of violence*, wherein a single agent invests in the ability to produce force. Second, there is a *balance of power*, which sustains peace by having all actors invest in the production of coercion.

We evaluate the ability to sustain order when agents do and do not have the ability to "exit" in the sense of Hirschman (1970). In the baseline analysis, we assume either agent can force the other to interact, either through violent conflict over wealth or, if not, through joint participation in an institution. In an extension we give both agents the ability to unilaterally escape interaction at a cost. We show that the costs of exit are inversely related to the ability to sustain order. As the cost of exit decreases, it becomes increasingly difficult to sustain political order.[3]

Besides characterizing the conditions for each type of political order to be sustainable as an equilibrium, we evaluate their relative efficiency. Several key results emerge. First, a peaceful state of nature, the first best outcome wherein peace prevails without any wasteful investment in coercion, can only exist under the implausible condition that conflict destroys the total value of the society's wealth. Second, when this condition is not met, a political institution that preserves peace through an investment in coercion exists only if the costs of conflict are large enough relative to the agents' uncertainty about the distributive consequences of conflict.

Third, we find that whenever structural conditions allow for a monopoly of violence it is also possible to sustain a balance of power with strictly lower investment in coercion. However, we also show that individual players can always obtain a larger payoff as a monopolist than they would get under any balance of power equillibrium. What is more, even within the set of supportable monopoly of violence equilibria, we find that the monopolist's payoff is greatest under a relatively inefficient institution. This occurs because the monopolist has an incentive to invest more in coercion than is strictly necessary to preserve peace, shrinking the total size of the pie but allowing her to obtain a larger portion in absolute terms. In this sense, not only are monopolies of violence generally inefficient, but the set of institutions we might expect if the strongest actors design them are inefficient relative to other supportable monopolies of violence. In this way we highlight how social efficiency and individual incentives cut hard against each other in the construction of social order.

---

[3]Substantively, this result comports empirical findings indicating that the relative appropriability of economic output is a crucial determinant of hierarchy (Allen 1997; Sanchez de la Sierra 2017; Scott 2017; Mayshar et al. 2018).

Finally, we explore institutional arrangements that do not fully maintain peace. To this end, we establish the conditions under which an inability to construct a peaceful political order results in conflict along the equilibrium path. When the costs of conflict are low or uncertainty is great, no peaceful institutional solution exists. As a consequence, political order cannot be assured. However, even when institutions that preserve the peace would be sustainable, they may not be economically efficient relative to ones that allow for a positive probability of violent conflict. The cost imposed by investing sufficiently in coercion to deter all violence may outweigh the costs of admitting occasional conflict.

Our approach to studying the construction of political order combines insight from theoretical literature that spans anthropology, political science, and economics. Broadly, existing theories understand the political order as an outcome of one of two social processes: cooperative bargaining or the coercive domination of some (typically the strong) upon others (typically the weak).

Proponents of *voluntaristic* theories conjecture that at some point in history, certain groups rationally and voluntarily constructed institutions to limit their behavior in order to purposefully achieve mutually beneficial outcomes. In their earliest form, theories of this sort take on a contractarian flavor.[4] In more recent incarnations, however, voluntaristic theories view institutions like the state as a response to market failures or collective action problems, arising deliberately to allow individuals and groups to coordinate their actions and achieve gains from cooperation (Childe 1946; Steward 1955; Gunawardana 1981; Ostrom 1990; Blanton and Fargher 2007). In the most famous of these contemporary voluntaristic theories, Wittfogel postulates the "hydraulic hypothesis" that states emerged when small communities abandoned individual autonomy to form a single political unit capable of coordinating large-scale irrigation projects (Wittfogel 1956, 1981). In other words, because of the economic gains that result from its presence, the state emerged functionally.

A second set of *conquest* theories treat political order as the outcome of violent conflict

[4]The most prominent example being Rousseau (2002)[1762].

between groups.[5] Rather than viewing political institutions as emerging explicitly to obtain economic gains, in these theories complex hierarchies comes into existence when those who are superior at producing violence enforce order through domination (Gumplowicz 1902; Oppenheimer 1922; Webster 1975; Naroll and Divale 1976; Cohen 1984). Here, any positive economic outcome that results from political order is ancillary to the conflictual processes that drive the state's construction. Classical sociologists like Oppenheimer, for example, assert that states came into existence when productive agriculturalists were conquered by nomadic pastoralists (Oppenheimer 1922, I I, pp. 51-55), a sentiment that is echoed by prominent political economy models (Olson 1993, 2000; Boix 2015).

Besides the standard critiques of functionalism, a clear problem with purely voluntaristic theories is that they disregard the violence that undergirds political order. And yet a purely coercive theory based upon the continued domination of one group over the other is similarly untenable. We rarely observe political order where violence is overt. Even in the most dictatorial environments, everyday coercion is latent; the application of violent force is unobserved. We combine features of both voluntaristic and coercive theories. Our approach allows us to know when actors, in the shadow of the threat of violence, can construct institutions that preserve peaceful order by assigning payoffs reflective of actors' abilities to coerce.

Existing formal models typically treat the construction of political order in one of two ways. The first fixes a game form and sees the emergence of state-like institutions as an equilibrium to this predefined game (Skaperdas 1992; Calvert 1995, 1998; Hirshleifer 1995; Hafer 2006; Piccione and Rubinstein 2007; Mayshar, Moav and Neeman 2017). The second approach takes a set of games, often one describing a state and another characterized as

[5]The earliest theories of this sort follow from Khaldūn (1958)[1377] and Bodin (1955)[1583]. Among modern scholars, Engels (2010)[1884], building on the anthropological work of Morgan (1907), was among the first to elucidate a conquest theory of state formation.

anarchy, and makes welfare comparisons between them, allowing a planner or decisive actor to choose between them (Moselle and Polak 2001; Grossman 2002; Konrad and Skaperdas 2012). Our approach combines the self-enforcing features of the "institutions as an equilibrium" approach with the understanding of institutions as formal rules as in the "institutions as constraints" approach. That is, we want to know when it is possible for agents in a state of nature to construct some rule that solves the problem of order. We then can make welfare comparisons between sets of feasible institutions.

A key facet of strategic interaction in anarchy motivating our approach is that individual actors may be unable to observe each other's coercive capabilities (Fearon 1995; Slantchev 2003). Mutual uncertainty makes it expensive to preserve peace. In particular, a peaceful institution must assure each player at least as much as she could expect from conflict if her privately known strength were as great as possible (Fey and Ramsay 2011, Result 3). Small groups with the ability to perfectly monitor each others' abilities and payoffs via informal social mechanisms may be able to preserve peace in the absence of the kind of formal political institutions we seek to understand.[6] For groups that cannot rely upon informal monitoring to reduce informational asymmetries, political institutions must endow each actor such that even the most powerful have no incentive to exercise their coercive advantage.

2 Model

The model consists of an interaction between two political actors representing individuals or self-organized political groups.[7] The actors may have an incentive to appropriate each other's wealth via violent conflict. We describe the conditions under which there is an institutional framework that averts conflict and characterize the institution that requires the

[6] See Ostrom (1990) on the centrality of monitoring mechanisms in obtaining cooperation.

[7] All of the substantive results of the analysis would hold in an environment with more than two players.

lowest investment in coercion.

There are two players, player 1 and player 2.[8] At the outset, each actor's share of the society's total wealth is $y_i > 0$, where $y_1 + y_2 = 1$. Before the two players interact with each other, each chooses an investment in coercive abilities m_i, where $0 \leq m_i \leq y_i$. Investing resources to produce coercive capacity increases a player's chances of winning in case conflict occurs but reduces the amount of wealth available for eventual consumption. That is, the more resources a player invests for violent purposes, the less she can devote towards productive ends. Let $M_i = [0, y_i]$ denote the set of feasible investment levels for each player.

In the model, the two players invest in force simultaneously.[9] After each player has chosen her own investment, m_i, she observes the other player's choice, m_j. At this point, each player simultaneously chooses whether to participate in a peaceful institution—the nature of which we describe below—or to opt for conflict instead. Let w_i denote each player's decision at this stage, where $w_i = 1$ represents conflict and $w_i = 0$ represents participation in the institution. Conflict occurs if either player chooses $w_i = 1$; the institution prevails only if $w_1 = w_2 = 0$. In this sense, we model institutions that are not only collectively beneficial, but give each player an individual incentive to opt for peace over conflict.

An institution in our model is simply a scheme for dividing wealth, depending on how much is left over after the players' coercive investments. In the model, an institution is defined as a pair of functions, $V_1(m_1, m_2)$ and $V_2(m_1, m_2)$, which represent how much wealth each player receives in the case where neither opts for conflict. We assume throughout that institutions are not wasteful, so $V_1(m_1, m_2) + V_2(m_1, m_2) = 1 - m_1 - m_2$.[10]

[8] We denote arbitrary players i and j.

[9] This need not be literally simultaneous; what is important is that neither actor can condition her investment on the other's investment.

[10] As our focus is on efficient institutions, our main substantive results would not change if we relaxed this assumption.

In this setting, all an institution does is redistribute wealth. This is, of course, a simplification—but one that sets a useful baseline for thinking about the conditions that enable actors in a state of nature to forego conflict.[11] Surprisingly, even the stronger actor, who would have an advantage in violent competition over society's wealth, will sometimes opt out of conflict in order to participate in a purely redistributive institution. By introducing additional benefits of institutions to the model, such as the reduction of transaction costs or the promotion of economic growth, we would simply expand the conditions under which peace is sustainable.

If at least one player instead opts for conflict, there is a violent contest over society's wealth. A player's investment in coercion, m_i, increases her chance of winning this struggle but reduces the prize—the amount of wealth that the winner receives. Each player's chance of winning, given the investment choices, is[12]

$$p_i(m_i, m_j) = \frac{\theta_i m_i}{\theta_i m_i + \theta_j m_j}. \tag{1}$$

The parameter $\theta_i > 0$ represents a player's *coercive effectiveness*: how much coercive capacity she can generate per unit of wealth she invests. The greater θ_i is, the cheaper it is for a player to build her capacity to a given level. For simplicity in the subsequent analysis, we label the players so that player 1 is the more effective one; i.e., we assume $\theta_1 \geq \theta_2$.

Even beyond the reduction in wealth due to the wasteful investment of resources to produce violence, conflict imposes costs on society. People are killed, fields are burned, and

[11] One might think of an institution as a more complicated set of rules that results in a distributive outcome. By arguments similar to those in the mechanism design literature (Myerson 1979; Fey and Ramsay 2009), our analysis identifies necessary structural conditions for any thicker institutional framework to produce peace.

[12] We may assume any distribution over victory in case $m_1 = m_2 = 0$, as the exact value of $p_i(0, 0)$ is inconsequential to the equilibrium analysis.

so on. In the model, the players know conflict is costly, but they only have partial information about how the costs will be distributed. To formalize this idea, let each player have a *type*, denoted t_i, that determines the actual distribution of costs. If $t_1 > t_2$, then player 1's costs are less than initially expected and player 2's are greater; the opposite is true if $t_1 < t_2$. We refer to a player's type as her privately known strength, or simply her strength; this is distinct from the coercive effectiveness parameters, θ_1 and θ_2, which are publicly known.

Each player has private information about her type.[13] Formally, let $F_i(t_i)$ denote the prior distribution of player i's type, which is common knowledge. Letting T_i denote the set of possible types (i.e., the support of the distribution), we assume T_i is bounded, with $\min T_i = \underline{t}_i$ and $\max T_i = \bar{t}_i$. The net cost player i bears for engaging in conflict is a function of both players' types, $c_i(t_i, t_j) = \bar{c}_i - t_i + t_j$, where $\bar{c}_i > 0$. Without loss of generality, we assume each t_i has mean zero,[14] so that $\bar{c}_i$ represents player i's *ex ante* expected cost.

Combining the mobilization-induced probabilities of victory and the type-dependent costs of fighting, a player's overall payoff from conflict is

$$W_i(m, t) = p_i(m_i, m_j)\,[1 - m_i - m_j] - c_i(t_i, t_j),$$

where $m = (m_1, m_2)$ and $t = (t_1, t_2)$ are the vectors of the players' mobilization choices and types, respectively. Because the players have private information about their types, a player may not know her exact payoff from conflict when she chooses whether to opt out of the institution. In this case, a player compares what the institution would give her to her

[13]Although it may be more natural to consider the coercive effectiveness parameters, θ_1 and θ_2, as private information, doing so significantly increases the technical challenge of the analysis without providing novel substantive insights. We therefore opt for the simpler model here.

[14]This assumption implies each $\underline{t}_i \leq 0$ and $\bar{t}_i \geq 0$. These inequalities hold strictly unless F_i places probability 1 on $t_i = 0$.

expected payoff from conflict:

$$\tilde{W}_i(m_i, m_j, t_i) = p_i(m_i, m_j)\,[1 - m_i - m_j] - \bar{c}_i + t_i - E[t_j \,|\, m_j]$$

This expected payoff is solely a function of the information available to a player at the time she chooses whether to opt out—her own type and both players' investments.

A player's expected utility from conflict depends on her type, but her payoff from the institution does not. This means we will focus on the incentives for the strongest type of a player, $\bar{t}_i$, to participate in the institution as opposed to engaging in conflict. Peace through an institution is sustainable as long as the strongest type prefers the institution over conflict, as then all weaker types have the same preference.

This is a multistage game of incomplete information, so we solve for perfect Bayesian equilibria (Fudenberg and Tirole 1991, 331–336). We consider various kinds of redistributive schemes, defined by the functions $V_1(m_1, m_2)$ and $V_2(m_1, m_2)$, in order to see what kinds of equilibrium behavior they may support. For our purposes, it does not matter whether the redistributive scheme arises endogenously from bargaining between the participants or whether it is proposed by an outside party. What matters is that the players have a shared expectation about how income will be distributed upon mutual participation in the institution.

We are particularly interested in *peaceful equilibria*, in which each player always opts to participate in the institution, and open conflict never occurs along the path of play. As it turns out, there are often numerous redistributive schemes that support peaceful equilibria. When this is the case, we look for those that do so with the least wasted wealth—i.e., the lowest total coercive investment, $m_1 + m_2$, along the path of play—and refer to them as *efficient* peaceful equilibria. Among peaceful equilibria, we only examine those in which all types of each player make the same coercive investment. This both simplifies the analysis and comports with our focus on efficiency: for any peaceful equilibrium sustained by varying coercive investments, there is a strictly more efficient peaceful equilibrium with a constant investment level.[15]

3 Peaceful Equilibria

There are three kinds of peaceful equilibrium. The first and simplest is a *peaceful state of nature*, in which peace is sustainable even though neither player invests in coercive capability. The peaceful state of nature represents circumstances under which the intrinsic incentives to engage in violence are too weak for conflict to be a concern.

The second type of peaceful equilibrium is a *monopoly of violence*, in which just one of the two players makes an investment in the production of violence. In this type of equilibrium, the monopolist invests enough in coercion to deter the other player from violent appropriation of wealth. Meanwhile, the institution is designed to ensure that the monopolist receives enough rents that she still prefers peace over the deployment of her coercive advantage. A monopoly of violence is sustainable under broader conditions than a peaceful state of nature; the greater the disparity in the players' coercive effectiveness or initial wealth, the broader these conditions are.

The last type of peaceful equilibrium is a *balance of power*, in which both players invest in the production of force and thereby deter each other from conflict. A balance of power may be sustainable when a monopoly of violence is not, particularly when the players are similar in coercive effectiveness. If the distribution of types is wide enough or the costs of conflict are low enough, even the balance of power may be unsustainable, meaning there is

[15]In a peaceful equilibrium, every type of a player must have the expected payoff, or else there would be an incentive for the types that receive less to mimic those that receive more (Fey and Ramsay 2011). From there, the inefficiency of a varying-investment equilibrium follows from the fact that each player's reservation value is strictly convex in the other's investment level, as we prove in Lemma 2 in the Appendix.

no institutional arrangement that assures peace. We consider the economic efficiency of each type of equilibrium—the level of coercive investment required to sustain them—and find, surprisingly, that for any monopoly of violence, there is a balance of power that is strictly less wasteful. However, if one player could unilaterally dictate the shape of political order, she would pick an inefficient monopoly of violence.

3.1 Peaceful State of Nature

We begin by characterizing the conditions under which peace is sustainable without any resources invested to produce force. As investment in coercion reduces the wealth available for the players to distribute, this is the least wasteful type of equilibrium. However, it is also the hardest to sustain. When one player does not invest, the other can gain an overwhelming advantage in conflict at a small cost.

For a player to prefer not to opt for violence, redistribution must give her at least as much as she expects from conflict. In terms of the model, then, a necessary condition for a peaceful state of nature is that each player receive at least her expected utility from conflict, given investments of $m_i = 0$ by both players:

$$V_i(0,0) \geq \tilde{W}_i(0,0,\bar{t}_i).$$

We state this condition for a player whose privately known strength, $\bar{t}_i$, is as large as possible, as that is the one with the greatest incentive for conflict.

While necessary, this condition is insufficient. Each player, expecting the other not to invest, may have an incentive to invest and then opt into a conflict that the other player did not prepare for. In order for a peaceful state of nature to be sustainable as an equilibrium, it must not be in either player's interest to deviate to making a small investment and forcing conflict. To formalize this idea, let a player's *reservation value*, denoted $\mathrm{RV}_i(t_i, m_j)$, be the greatest expected utility she can attain by investing and forcing conflict, given her own type and how much she expects the other player to invest:

$$\mathrm{RV}_i(t_i, m_j) = \sup_{m_i \in M_i} \tilde{W}_i(m_i, m_j, t_i).$$

When one player expects the other to invest nothing, as in a peaceful state of nature, she can assure herself victory in conflict with any $m_i > 0$, even a very small one. Therefore, the reservation value of a player who expects no investment by the other is simply

$$\mathrm{RV}_i(t_i, 0) = 1 - \bar{c}_i + t_i.$$

Because the strongest type of each player is the one with the greatest incentive to deviate to conflict, a sufficient condition for each player to participate in a peaceful state of nature is

$$V_i(0, 0) \geq \mathrm{RV}_i(\bar{t}_i, 0). \tag{2}$$

If this condition holds for each player, neither has an incentive to take advantage of the other by investing and fighting.

Under what conditions does peace prevail in the state of nature? If neither player invests, then the redistributive scheme divides all of their initial wealth: $V_1(0,0)+V_2(0,0) = y_1+y_2 = 1$. The critical question, then, is whether the unit of wealth is enough to distribute between the players while preserving peace—i.e., that the no-deviation condition, Equation 2, can be met for each player. Formally, there is enough wealth to satisfy the strongest type of each player only if $\mathrm{RV}_1(\bar{t}_1, 0) + \mathrm{RV}_2(\bar{t}_2, 0) \leq 1$, which is equivalent to

$$\bar{c}_1 + \bar{c}_2 \geq 1 + \bar{t}_1 + \bar{t}_2.$$

As we summarize in the following proposition, this condition fully determines whether there is a peaceful state of nature.[16]

Proposition 1. *There is an equilibrium with a peaceful state of nature if and only if* $\bar{c}_1 + \bar{c}_2 \geq 1 + \bar{t}_1 + \bar{t}_2$.

Evidently, it is quite difficult to sustain peace in the absence of coercive investments. Specifically, the expected costs of conflict must exceed the total value of the society's wealth. Even with such high costs of conflict, a peaceful state of nature may still be unsustainable if there is enough uncertainty about the distribution of the costs of conflict, represented here by the upper bound on players' privately known strength, $\bar{t}_i$. The logic of this result is closely connected to the seminal finding in the international conflict literature that incomplete information is a cause of war (Fearon 1995).

3.2 Monopoly of Violence

If the expected cost of conflict is too low or the uncertainty about players' types is too great, then we cannot expect peace to prevail in the absence of organized force. We now consider the sustainability and efficiency of political arrangements in which one player maintains peace by establishing a monopoly over the use of coercive force.

In a monopoly of violence, one player (call her the monopolist) invests $m_i > 0$ along the path of play, thereby reducing the incentive of the other player (the subject) to opt for conflict. Meanwhile, the monopolist's temptation to opt for violence over peace, given her coercive advantage, can be restrained as long as the distribution of wealth in case of peace is sufficiently favorable to her. In other words, in a monopoly of violence, the monopolist collects rents from the subject as the price of preserving the peace.

In determining whether a monopoly of violence may produce peace, we run into a fundamental strategic tension. On one hand, the monopolist must invest enough to deter the subject from partaking in violence. To formalize the idea here, consider an equilibrium in

[16] All proofs are in the Appendix.

which player i is the monopolist and invests $m_i^* > 0$, while player j is the subject and does not invest in coercion. Suppose the equilibrium gives V_i^* to player i and V_j^* to player j, where $V_i^* + V_j^* = 1 - m_i^*$. The equilibrium must give the strongest type of the subject as much as she could expect from optimal coercive investment:

$$V_j^* \geq \mathrm{RV}_j(\bar{t}_j, m_i^*).$$

Because each player's reservation value is decreasing in the other's investment, it becomes easier for this condition to hold as m_i^* increases. The more the monopolist invests, the more likely the subject is to be deterred.

On the other hand, the more the monopolist invests to deter the subject, the less wealth there is left over to be distributed peacefully. The greater the cost of deterrence, the harder it becomes to design an institution that gives the monopolist an incentive to participate. The temptation for the monopolist is to deviate to investing an infinitesimal amount, which is still enough to assure victory over a subject who spends nothing and leaves more wealth than if the monopolist invests enough to deter. Formally, the condition for the monopolist always to prefer the equilibrium distribution of wealth over opting out is $V_i^* \geq \mathrm{RV}_i(\bar{t}_i, 0)$, which is equivalent to

$$1 - m_i^* - V_j^* \geq 1 - \bar{c}_i + \bar{t}_i.$$

It becomes harder for this condition to hold as m_i^* increases, as the cost of deterrence is eventually unbearable.

In summary, the basic strategic tension is that the monopolist's investment must be great enough to deter the subject, but not so great that she would rather fight over a larger pie. The formal condition is that there exist $m_i^* > 0$ such that

$$\mathrm{RV}_i(\bar{t}_i, 0) + \mathrm{RV}_j(\bar{t}_j, m_i^*) \leq 1 - m_i^*.$$

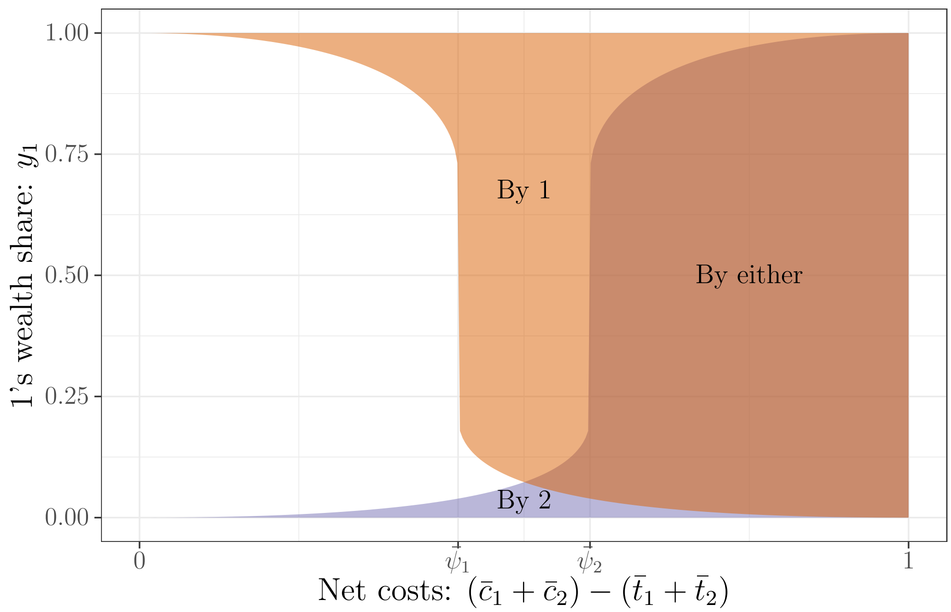


Figure 1. Existence of a monopoly of violence as a function of the expected net cost of conflict and the stronger player's share of the initial wealth.

Assuming this condition can be met at all—i.e., that it is possible to deter the subject while leaving enough rents for the monopolist to extract—our goal is to find the lowest level of investment m_i^* at which it does. This represents the least wasteful, or most economically efficient, monopoly of violence.

Two factors determine whether a monopoly of violence can sustain peace. The first is the expected total cost of conflict. The more costly conflict is, the less one must invest to deter the other player from conflict and thus the easier it is to sustain peace through a monopoly of violence. The second is the distribution of initial wealth, which can cut either way. Even if the costs of conflict are relatively high, a monopoly of violence may be unsustainable if the prospective monopolist does not have sufficient initial wealth to make the necessary investment in coercion. By the same token, a player with an inordinate share of the initial wealth may be able to sustain a monopoly of violence even when the expected costs of conflict are low, simply because the other player lacks the capacity to resist.

In summary, a monopoly of violence requires that the costs of conflict be high, that the initial distribution of wealth be skewed heavily in favor of the monopolist, or both. Figure 1 illustrates these conditions, and the following proposition states them formally.

Proposition 2. *There are cost thresholds for a monopoly of violence, $\bar{\psi}_1$ and $\bar{\psi}_2$, such that:*

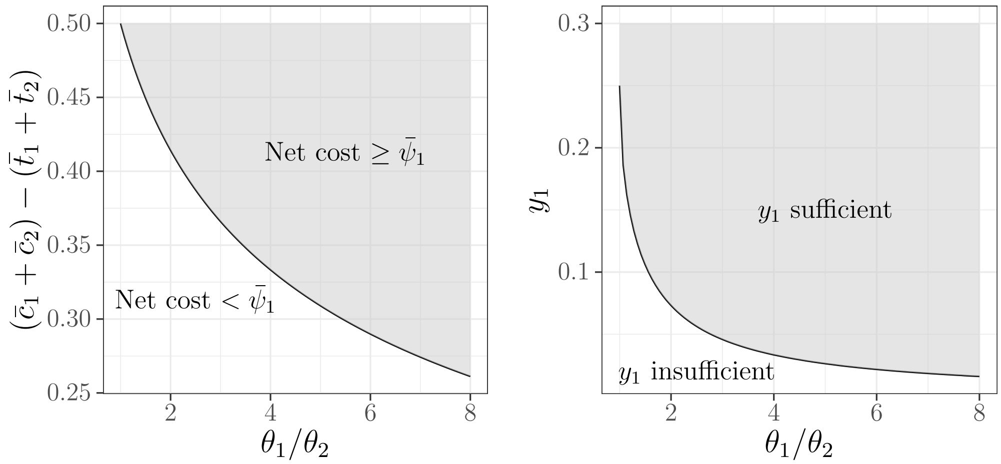
(a) $0 < \bar{\psi}_1 \leq 1/2$ *and* $\bar{\psi}_2 = 1 - \bar{\psi}_1 \geq 1/2$.

(b) If $\bar{\psi}_i + \bar{t}_i + \bar{t}_j \leq \bar{c}_i + \bar{c}_j < 1 + \bar{t}_i + \bar{t}_j$, then there is a peaceful equilibrium with a monopoly of violence by player i if and only if initial wealth is not too skewed in favor of player j.

(c) If $\bar{t}_i + \bar{t}_j < \bar{c}_i + \bar{c}_j < \bar{\psi}_i + \bar{t}_i + \bar{t}_j$, then there is a peaceful equilibrium with a monopoly of violence by player i if and only if initial wealth is skewed far enough in favor of player i.

In a monopoly of violence equilibrium, the equilibrium level of investment by the monopolist must be enough to deter the subject from forcing a conflict. The greater the monopolist's coercive advantage over the subject, the cheaper it is to do so. This line of logic leads us to two conclusions about peaceful equilibria with a monopoly of violence. First, it is easier to sustain an equilibrium with the player whose coercive effectiveness is greater (which we have labeled as player 1) as the monopolist. Second, the greater the imbalance in coercive effectiveness, the easier it is to support a monopoly of violence in the first place. A peaceful monopoly of violence is hardest to establish when the players have equal abilities to translate investment into coercive capacity. An imbalance in coercive effectiveness decreases the cost of sustaining a monopoly of violence, and with it the constraint on the monopolist's initial wealth, as illustrated in Figure 2.

If the costs of conflict are large enough relative to the magnitude of the players' uncertainty, then a monopoly of violence by either player is potentially sustainable as an equilibrium. It is less wasteful to have the player with greater coercive effectiveness be the monopolist, as the other player can be deterred with less effort, but this might be impossible if the initial distribution of wealth is skewed against the more effective player. If the less



(a) Cost condition. **(b)** Wealth condition.

Figure 2. Conditions for the existence of a monopoly of violence as a function of the stronger player's coercive advantage.

effective player disproportionately controls the initial wealth and the costs of conflict are close enough to the threshold defined in Proposition 2, then there is no peaceful equilibrium with a monopoly of violence.

Given the opportunity, the player with greater coercive effectiveness would indeed choose to be the monopolist. However, this does not necessarily mean she would choose to invest at the socially efficient level. In fact, we find that the equilibrium with the highest payoff for the monopolist entails strictly more investment than is necessary.

Proposition 3. *In a monopoly of violence, the monopolist prefers more coercive investment than the socially efficient level.*

This result may seem counterintuitive, as an over-investment in coercion shrinks the size of the pie that is redistributed in a peaceful equilibrium. But shrinking the pie, up to a certain point, is strategically advantageous for the monopolist. The less wealth there is left over after coercive investment, the less incentive the subject has to engage in costly conflict over that wealth. Consequently, the subject's reservation value shrinks rapidly with the monopolist's investment, allowing the monopolist to extract more from redistribution while maintaining the peace. At low levels, the marginal reduction in the subject's reservation value due to the monopolist's investment outweighs the marginal effect on the size of the pie, giving the monopolist an incentive to over-invest.

3.3 Balance of Power

We now consider the final type of peaceful equilibrium, which we term a balance of power, in which each player invests in coercion to deter the other from an attempt to violently appropriate wealth. It is easier to meet the conditions for a balance of power equilibrium than for a monopoly of violence—whenever a monopoly of violence is a sustainable, so too is a balance of power, but the reverse is not true. More interestingly, for any monopoly of violence, there is a balance of power that attains peace at strictly lower cost. The efficiency advantage of a balance of power is most pronounced when the two players' coercive effectiveness is roughly equal.

In a balance of power equilibrium, player 1 invests $m_1^* > 0$, player 2 invests $m_2^* > 0$, and each opts for the institution over conflict. As before, in order for the strongest type of each player, $\bar{t}_i$, to prefer redistribution over conflict, it is necessary but insufficient that her promised portion equal at least what she would get from fighting:

$$V_i(m_i^*, m_j^*) \geq \tilde{W}_i(m_i^*, m_j^*, \bar{t}_i).$$

If the strongest type expects conflict, she may prefer to invest more or less than the amount necessary to deter the other player, given her expectation that the other player will invest m_j^*. Therefore, peace requires that the redistributive scheme give the strongest type of each player at least what she would expect from optimal investment in anticipation of conflict:

$$V_i(m_i^*, m_j^*) \geq \mathrm{RV}_i(\bar{t}_i, m_j^*).$$

Because this condition must hold for both players, a balance of power equilibrium requires that

$$\mathrm{RV}_1(\bar{t}_1, m_2^*) + \mathrm{RV}_2(\bar{t}_2, m_1^*) \leq V_1(m_1^*, m_2^*) + V_2(m_1^*, m_2^*) = 1 - m_1^* - m_2^*.$$

The critical question is whether there is a pair of investments for which this condition holds. If not—and if the conditions for a peaceful state of nature and a monopoly of violence do not hold either—then there is no peaceful arrangement of political order.

In a balance of power equilibrium, each player must invest enough to deter the other. The greater the expected cost of conflict, the cheaper it is to do so. Consequently, the main condition for a balance of power equilibrium is that the expected cost of conflict be great enough. However, unlike with a monopoly of violence, the players' relative coercive effectiveness and initial wealth do not affect the sustainability of peace through a balance of power. As one player's coercive advantage increases, the cost of deterring that player from conflict increases at the same rate as the cost of deterring the other one decreases. Because the effects cancel each other out, the cost condition for a balance of power equilibrium is independent of relative coercive effectiveness, and there is no constraint on initial wealth.

Proposition 4. *There is a peaceful equilibrium with a balance of power if and only if* $\bar{c}_1 + \bar{c}_2 \geq \bar{t}_1 + \bar{t}_2$.

Naturally, this is weaker than the cost conditions for a peaceful state of nature or monopoly of violence. Moreover, because each $\bar{c}_i > 0$, this condition is sure to hold if there is little uncertainty about the distribution of the costs of conflict (i.e., each $\bar{t}_i \approx 0$). On the other hand, if uncertainty is great enough, even a balance of power cannot sustain peace, and any equilibrium of the game entails a positive probability of violent conflict.

3.4 Comparing Political Orders

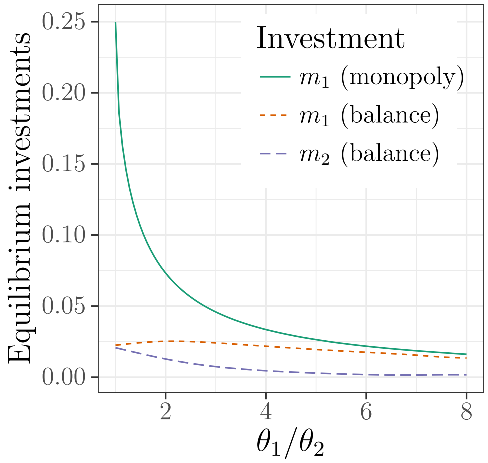
Having characterized each potential arrangement of political order, we now consider social welfare. Which political arrangement obtains peace at the lowest cost? An equilibrium with investment levels (m_1^*, m_2^*) results in a final total wealth of $1 - m_1^* - m_2^*$, so the question is effectively which kind of equilibrium attains peace at the lowest value of $m_1^* + m_2^*$. Obviously, if the conditions of Proposition 1 are met and there is a peaceful state of nature, this is the most economically efficient equilibrium. Short of that, we find that a balance of power can always obtain peace at a lower cost than any monopoly of violence. In other words, a monopoly of violence is never the most efficient form of social order in our model.

Proposition 5. *If there is a peaceful equilibrium with a monopoly of violence, then there is a peaceful equilibrium with a balance of power with strictly less total coercive investment.*

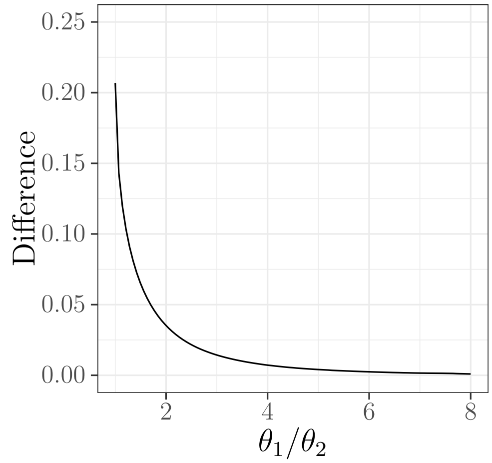
This result follows from the decreasing returns to investment as an instrument of deterrence. The first unit of investment does much more to decrease a player's reservation value than does the second, which in turn does more than the third. When a peaceful state of nature is not sustainable, the problem at hand is to identify the equilibrium that lowers the total reservation value just enough that the leftover wealth can make each player prefer the institution over conflict. This can be accomplished more cheaply by having both players spend a bit than by having a single player spend a lot. In other words, if we took any monopoly of violence, had the monopolist invest a bit less and the subject invest a bit more, we could still sustain peace with room to spare.

This result emerges in part from our technological assumptions about the relationship between investment and coercive power—i.e., the shape of the function $p_i(m_i, m_j)$, defined in Equation 1. If there were substantial economies of scale in the production of coercive force, then it might be more efficient to have one player make the entire investment. The result also depends on the fact that institutions merely redistribute wealth. If institutions provided public goods besides the prevention of violence and there were economies of scale in the production of public goods, then a monopolist of violence might be more efficient. Nonetheless, we find it surprising and instructive that a monopoly of violence is always inefficient in the baseline setting we construct.

The magnitude of the inefficiency in a monopoly of violence depends on how imbalanced the players are in their coercive effectiveness. The closer they are to equality, the less



(a) Least costly equilibrium.



(b) Cost difference between monopoly of violence and balance of power.

Figure 3. Illustration of the efficient equilibrium and the efficiency gap as a function of the power difference.

inefficient a monopoly of violence will be, as illustrated in Figure 3. At parity, it requires a substantially larger investment to maintain a monopoly of violence than the most efficient balance of power. However, as the coercive advantage of the stronger player grows, the equilibrium investment of the weaker player in the most efficient balance of power shrinks. Consequently, the efficiency difference between this and the least expensive monopoly of violence becomes negligible.

If monopolies of violence are inefficient, and balances of power are sustainable as equilibria, why should we ever observe the monopolization of force by a sovereign government? The problem is that the best equilibrium for the society as a whole is not necessarily the best for the monopolist. If one player could dictate the choice of equilibrium, she would select a monopoly of violence. The following results extends Proposition 3 by showing that not only is the best monopoly of violence for a player one in which she over-invests, but that the player prefers this monopoly over any balanced political order.

Proposition 6. *The best peaceful equilibrium for player i is a monopoly of violence by i with more coercive investment than is socially efficient, if such an equilibrium exists.*

In summary, we have characterized the conditions for each type of peaceful equilibrium and uncovered some important implications for the efficiency of political orders. When some level of coercive investment is necessary to preserve peace, the cheapest way to do so involves investment by both players, with the majority coming from the player with the greater coercive effectiveness. Precisely because a balance of power is cheaper to sustain, it is supportable under a wider set of conditions on the expected costs of conflict and the initial distribution of wealth than is a monopoly of violence. However, social efficiency and individual incentives do not coincide. If a single player could dictate the nature of political order, she would choose a monopoly of violence in which she is the monopolist and invests more than is necessary to deter the other player from conflict.

4 Equilibria with Conflict

So far we have focused on peaceful equilibria where the threat of conflict may shape redistributive outcomes but conflict never occurs along the equilibrium path. The conditions for assured peace are quite stringent. In particular, it must be possible to identify an institution that satisfies the strongest type of each player. If there is even a small probability that one player is extraordinarily strong, then this condition becomes impossible to meet, meaning there is no equilibrium that always ends peacefully. In this case, the most efficient political order involves a positive probability of violence.

To illustrate efficient political orders with a positive probability of violence, we consider a simple case of the model with one-sided private information.[17] Let the players have equal initial wealth ($y_1 = y_2 = 1/2$), coercive effectiveness ($\theta_1 = \theta_2 = \theta$), and expected costs of conflict ($\bar{c}_1 = \bar{c}_2 = \bar{c}$). Moreover, let it be common knowledge that $t_2 = 0$, and let player 1's type be drawn from $T_1 = \{0, \bar{c}\}$. Substantively, this means that the total cost of conflict is

[17] Our substantive results would be qualitatively similar with two-sided private information or asymmetries between the players.

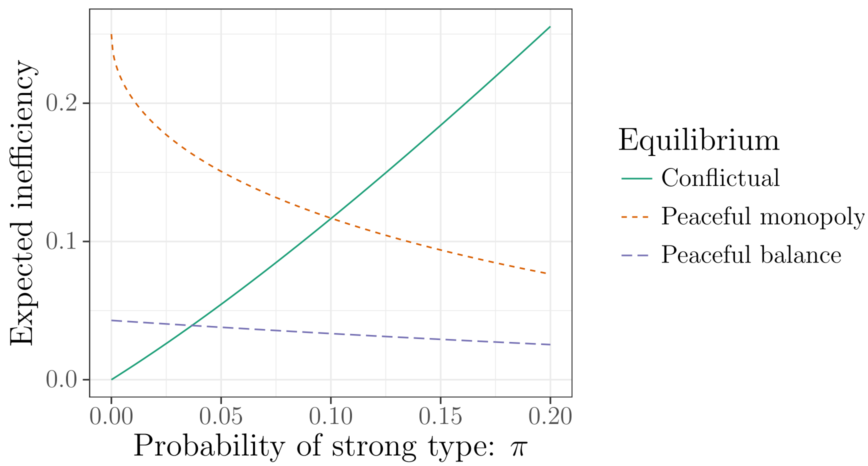


Figure 4. Expected inefficiency of the proposed equilibrium with a positive probability of conflict, compared to the most efficient peaceful monopoly of violence and balance of power.

$2\bar{c}$, which will either be evenly divided or fall exclusively on player 2; initially only player 1 knows which is the case.[18] Let π denote the prior probability that player 1 is strong, i.e., that $t_1 = \bar{c}$, where $0 < \pi < 1$.

The problem with a peaceful equilibrium in this setting is that the players must spend an inordinate amount of their wealth to deter the strong type from conflict, even if the probability of such a type is small. It would require less investment to deter only the weak type and plan for conflict with the strong type. To formalize this idea, imagine a conditional monopoly of violence, in which player 2 always invests, while only the strong type of player 1 invests. If player 1 is the low type, player 2 acts as the monopolist; otherwise, conflict occurs. If the costs of conflict are great enough, this conditional monopoly of violence is sustainable as an equilibrium.[19] More importantly, when the prior probability of a strong type is low enough, the expected efficiency loss in the conditional monopoly of violence is less than that of any peaceful equilibrium. Figure 4 illustrates this result, showing that the conditional monopoly approaches perfect efficiency as the prior probability of a strong type goes to zero.

[18]Implicitly this relaxes the assumption that $E[t_1] = 0$. This allows us to take comparative statics on the prior probability of a strong type while holding the type space fixed.

[19]Proposition 7 in the Appendix provides a formal characterization.

There are two sources of economic inefficiency in the conditional monopoly of violence. The first is the amount that player 2 and the strong type of player 1 invest in their coercive capacity. Intuitively, however, we expect these to be less than in the baseline peaceful equilibrium. The other, more important source of inefficiency is the cost of conflict, which is now realized on the path of play. In the equilibrium proposed here, conflict occurs whenever player 1 is strong. Consequently, if the prior probability of such a type, π, is large enough, the efficiency gains of lower coercive investment are swamped by the efficiency loss due to conflict. By the same token, when π is small enough, so too is the efficiency loss from the costs of conflict.

When the prior probability of a strong type is low enough, the equilibrium we construct with a positive probability of conflict is less wasteful in expectation than the best peaceful monopoly of violence or balance of power. Although conflict remains *ex post* inefficient, it can be efficient *ex ante* to allow for some chance of conflict, so as to reduce the extreme cost of guaranteeing participation in the institution.[20]

5 Exit

We now extend the model to allow players to opt out of interacting with each other altogether, whether peacefully or violently. In the *game with exit*, when the players have learned their types and are choosing how much to mobilize, each may instead choose to exit the interaction. If either player chooses to exit, then there is no further interaction, and each player consumes a fraction of her initial wealth. Otherwise, the game proceeds as in the original model.

[20]This result is connected to work in international relations theory demonstrating that the high cost of arming may be a cause of interstate war (Coe 2011). It is also closely related to the well-known finding in economics that there are generally not *ex post* efficient trading mechanisms that are compatible with individual incentives (Myerson and Satterthwaite 1983).

We make three assumptions about the value of exiting the interaction. First, the greater a player's initial wealth, the more attractive the exit option is. An initially wealthier player has more to lose by interacting, whether in peaceful redistribution or violent conflict. Second, exit is economically inefficient. Whether due to returns to scale, complementarities in production, or gains from specialization, the players' resources can produce more when combined than when apart. Specifically, we assume that the most a player can receive from exit is αy_i, where $0 < \alpha < 1$. Third, the incentive to exit is greater for stronger types of a player. In the original model, a player's type represents her privately known ability to mitigate the costs of violent conflict; it is natural to assume that the same traits also determine a player's ability to thrive under anarchy. To incorporate this assumption into the model, we assume the cost of exit to type t_i of player i is $\beta(\bar{t}_i - t_i)$, where $\beta \geq 0$.[21]

Let $e_i(t_i)$ denote the payoff from exiting to type t_i of player i. The above assumptions imply

$$e_i(t_i) = \alpha y_i - \beta(\bar{t}_i - t_i).$$

In case player i invests $m_i > 0$ and player j chooses to exit, we assume player i's coercive investment is wasted, so player i receives $\alpha(y_i - m_i) - \beta(\bar{t}_i - t_i)$. We now consider how exit alters the sustainability and shape of a monopoly of violence when the monopolist has a temptation to exit.

Naturally, the conditions to support a monopoly of violence become more stringent once we introduce the possibility of exit. In the baseline model, an imbalance of initial wealth does not threaten a monopoly of violence as long as the imbalance favors the monopolist. With the possibility of exit, however, high initial wealth may induce the potential monopolist to exit rather than to mobilize and expose her wealth to redistribution. Therefore, all else

[21] The results of the extension would be substantively the same if the cost of exit were any decreasing function of the player's type, and if the cost of exit for the strongest type were nonzero.

equal, a balanced distribution of wealth is most conducive to a monopoly of violence in the game with exit.

Even when a monopoly of violence remains sustainable, it may require more coercive investment, making it less economically efficient, than in the baseline game. In the baseline model, the most efficient monopoly of violence entails the player with greater coercive effectiveness (namely, player 1) spending the least possible to deter the other player from conflict while leaving enough surplus for the monopolist. If exit is attractive enough for the monopolist, then the redistributive surplus required to induce her to participate increases; this in turn requires her to invest more to make conflict less attractive for the subject.[22]

When a full monopoly of violence is unsustainable, there may be an equilibrium with partial exit. In such an equilibrium, when the monopolist controls a disproportionate amount of the initial wealth, stronger types of the monopolist exit, weaker types of the monopolist enter and mobilize enough to deter the subject, and the subject always enters. Such an equilibrium requires that the exit payoff be low enough for weaker types of the monopolist, i.e., that β be large enough.

Partial exit by the monopolist has a complicated effect on the subject's incentive to deviate to conflict. On one hand, the direct incentive to invest decreases with the probability of exit by the monopolist. If the subject deviates to invest $m_j > 0$ and the monopolist does not enter, then the subject has effectively wasted a portion of her wealth. As the monopolist becomes overwhelmingly likely to exit, the subject's *ex ante* incentive to invest in coercive capacity vanishes.

On the other hand, the more likely the monopolist is to exit, the greater is the subject's payoff from conflict conditional on entry occurring. In an equilibrium with partial exit, the subject infers from failure to exit that the monopolist is relatively weak. This increases the subject's expected utility from conflict and thus her reservation value. Substantively, this means the subject can more credibly threaten to reject the redistributive scheme when the

[22]See Proposition 8 in the Appendix.

probability of exit is high, because her coercive position is relatively strong if an interaction takes place.

The upshot is that partial exit by the monopolist has ambiguous distributive consequences. If the probability of exit is small, then the subject might be more tempted than in the baseline game to invest and force a conflict conditional on entry, knowing that the monopolist is relatively weak. In this case, the redistributive scheme in an equilibrium with partial exit might have to provide more to the subject than is required in the baseline game. However, if the probability of exit is large enough, the temptation to invest and force conflict with entrants disappears. Once the subject has chosen not to mobilize, she is sure to lose any conflict with the monopolist. So, even knowing that the monopolist is weak will not tempt her to engage in conflict *ex post*. In this case, then, redistribution in an equilibrium with partial exit may be unfavorable to the subject, relative to the baseline game.

6 Conclusion

When is peaceful political order self-enforcing? Under what conditions can monopolies of violence be sustained? If these conditions are met, what are the welfare implications of order? The model we developed in this paper answers these questions, describing the necessary conditions to construct political order. What is more, we have shown that orders characterized by a monopoly of force are generally inefficient relative to political orders where multiple agents maintain coercive abilities. Furthermore, even within the set of peace-preserving institutions backed by a monopoly of violence, the institution most preferred by the monopolist requires an inefficiently high investment in coercion.

These results provide a potential political explanation for observed lower levels of welfare in the earliest of states, a phenomenon that scholars working in anthropological traditions have heretofore related to negative epidemiological or ecological externalities associated with the formation of states. Besides informing our understanding the distant history of *de novo* state formation, our model allows us to better contextualize the state and the system of states we observe in the contemporary world.

We suggest that organizing principles defined by sovereign constituent units that monopolize violence are not "natural" in the way many contemporary observers of international relations might assert. Indeed, political order based upon diffuse coercive abilities is, as we have shown, sustainable whenever there is peace based upon a monopolist of force. More surprisingly, we find that this diffuse balance of power is welfare-improving.

Why then does the state persist? Consider two plausible answers. First, it could be that norms of mutual recognition exclude non-states from the international system. While potentially true, our model suggests a second likely answer. Monopolies of violence, though inefficient, persist because they endow the most powerful actors with the greatest payoff. If the powerful are capable of establishing the rules of the game, we expect them to select socially inefficient yet individually optimal institutional arrangements.

References

Allen, Robert C. 1997. "Agriculture and the Origins of the State in Ancient Egypt." *Explorations in Economic History* 34(2):135–154.

Bates, Robert H. 2008. "State Failure." *Annual Review of Political Science* 11.

Besley, Timothy and Torsten Persson. 2011. *Pillars of prosperity: The political economics of development clusters.* Princeton University Press.

Blanton, Richard and Lane Fargher. 2007. *Collective action in the formation of pre-modern states.* Springer Science & Business Media.

Bodin, Jean. 1955. *Six books of the commonwealth.* Vol. 5-6 B. Blackwell Oxford.

Boix, Carles. 2015. *Political Order and Inequality: Their Foundations and their Consequences for Human Welfare.* Cambridge University Press.

Calvert, Randall. 1995. "The rational choice theory of social institutions: Cooperation, coordination, and communication." *Modern political economy: Old topics, new directions* pp. 216–268.

Calvert, Randall L. 1998. "Explaining Social Order: Internalization, External Enforcement, or Equilibrium?" *Institutions and social order. The University of Michigan Press, Ann Arbor* pp. 131–161.

Childe, Vere Gordon. 1946. *What happened in history*. Penguin books New York.

Coe, Andrew J. 2011. "Costly peace: A new rationalist explanation for war." *Working Paper, University of Southern California* `https://www.andrewjcoe.com/research`.

Cohen, Mark Nathan. 1989. *Health and the Rise of Civilization*. Yale University Press.

Cohen, Ronald. 1984. Warfare and State Formation: Wars Make States and States Make Wars. In *Warfare, Culture, and Environment*, ed. Brian R Fergeson. Academic Press, Orlando pp. 329–358.

Edgerton, Robert B. 2010. *Sick societies*. Simon and Schuster.

Engels, Friedrich. 2010. *The origin of the family, private property and the state*. Penguin UK.

Fearon, James D. 1995. "Rationalist explanations for war." *International organization* 49(03):379–414.

Fey, Mark and Kristopher W Ramsay. 2009. "Mechanism design goes to war: peaceful outcomes with interdependent and correlated types." *Review of Economic Design* 13(3):233–250.

Fey, Mark and Kristopher W Ramsay. 2011. "Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict." *American Journal of Political Science* 55(1):149–169.

Fudenberg, Drew and Jean Tirole. 1991. *Game Theory*. Cambridge, Mass.: MIT Press.

Grossman, Herschel I. 2002. "Make us a king: anarchy, predation, and the state." *European Journal of Political Economy* 18(1):31–46.

Gumplowicz, Ludwig. 1902. *Die Sociologische Staatsidee*. Wagner.

Gunawardana, Ralph. 1981. Social function and political power: a case study of state formation in irrigation society. In *The Study of the State*, ed. Henri J.M. Claessen and Peter Skalnik. The Hauge: Moulton Publishers chapter 7, pp. 133–154.

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

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard University Press.

Hirshleifer, Jack. 1995. "Anarchy and its breakdown." *Journal of Political Economy* pp. 26–52.

Khaldūn, Ibn. 1958. *The Muqaddimah: An Introduction to History In Three Volumes*. Princeton University Press.

Konrad, Kai A and Stergios Skaperdas. 2012. "The Market for Protection and the Origin of the State." *Economic Theory* 50(2):417–443.

Larsen, Clark Spencer. 1995. "Biological changes in human populations with agriculture." *Annual Review of Anthropology* 24(1):185–213.

Levy, Brian and Sahr Kpundeh. 2004. "Building State Capacity in Africa: New Approaches, Emerging Lessons." *World Bank Publications* .

Mayshar, Joram, Omer Moav and Zvika Neeman. 2017. "Geography, Transparency, and Institutions." *American Political Science Review* 111(3):622–636.

Mayshar, Joram, Omer Moav, Zvika Neeman and Luigi Pascali. 2018. “The Emergence of Hierarchies and States: Productivity vs. Appropriability.” *Working Paper, University of Warwick* `https://warwick.ac.uk/fac/soc/economics/staff/omoav/mmnp11aug2017.pdf`.

Morgan, Lewis Henry. 1907. *Ancient society; or, researches in the lines of human progress from savagery, through barbarism to civilization*. H. Holt.

Moselle, Boaz and Benjamin Polak. 2001. “A model of a predatory state.” *Journal of Law, Economics, and Organization* 17(1):1–33.

Myerson, Roger B. 1979. “Incentive compatibility and the bargaining problem.” *Econometrica: journal of the Econometric Society* pp. 61–73.

Myerson, Roger B and Mark A Satterthwaite. 1983. “Efficient mechanisms for bilateral trading.” *Journal of economic theory* 29(2):265–281.

Naroll, Raoul and William Tulio Divale. 1976. “Natural selection in cultural evolution: warfare versus peaceful diffusion.” *American Ethnologist* 3(1):97–129.

North, D.C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.

Olson, Mancur. 1993. “Dictatorship, democracy, and development.” *American Political Science Review* pp. 567–576.

Olson, Mancur. 2000. *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships*. Basic books.

Oppenheimer, Franz. 1922. *The state: Its history and development viewed sociologically*. Vanguard Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.

Piccione, Michele and Ariel Rubinstein. 2007. "Equilibrium in the Jungle*." *The Economic Journal* 117(522):883–896.

Rotberg, Robert I. 2002. "The new nature of nation-state failure." *Washington Quarterly* 25(3):83–96.

Rousseau, Jean-Jacques. 2002. *The social contract: And, the first and second discourses.* Yale University Press.

Sanchez de la Sierra, Raul. 2017. "On the origin of states: Stationary bandits and taxation in Eastern Congo." *Working Paper, University of California, Berkeley* `https://raulsanchezdelasierra.files.wordpress.com/2013/09/revised-manuscript.pdf`.

Sandeford, David S. 2018. "Organizational complexity and demographic scale in primary states." *Royal Society open science* 5(5):171137.

Scott, James C. 2017. *Against the Grain: A Deep History of the Earliest States.* Yale University Press.

Skaperdas, Stergios. 1992. "Cooperation, conflict, and power in the absence of property rights." *The American Economic Review* pp. 720–739.

Slantchev, Branislav L. 2003. "The principle of convergence in wartime negotiations." *American Political Science Review* 97(04):621–632.

Steward, Julian Haynes. 1955. *Irrigation Civilizations: A Comparative Study. A Symposium on Method and Result in Cross Cultural Regularities.* Vol. 1 Pan American Union.

Webster, David. 1975. "Warfare and the Evolution of the State: A Reconsideration." *American Antiquity* pp. 464–470.

Wittfogel, Karl August. 1956. *The hydraulic civilizations.* University of Chicago Press.

Wittfogel, Karl August. 1981. *Oriental despotism: A comparative study of total power*. Vol. 3 Vintage books New York.

A Supplemental Appendix to "The Construction of Political Order"

Contents

A.1 Proof of Proposition 1

Proposition 1. *There is an equilibrium with a peaceful state of nature if and only if* $\bar{c}_1 + \bar{c}_2 \geq 1 + \bar{t}_1 + \bar{t}_2$.

Proof. The argument in the text proves the condition is necessary. For sufficiency, we assume the condition holds and construct an equilibrium. Let $V_1(0,0) = \mathrm{RV}_1(\bar{t}_1, 0)$, so that $V_2(0,0) = 1 - \mathrm{RV}_1(\bar{t}_1, 0)$. For all $(m_1, m_2) \neq (0,0)$, let each $V_i(m_1, m_2) = (1 - m_1 - m_2)V_i(0,0)$, so that $V_1(m_1, m_2) + V_2(m_1, m_2) = 1 - m_1 - m_2$ as required. We claim that the following assessment constitutes a peaceful state of nature equilibrium:

- Every type of each player chooses $m_i = 0$.
- After observing any m_j, player i's updated belief about t_j equals her prior.
- After mobilization choices (m_i, m_j), type t_i of player i chooses $w_i = 0$ if $V_i(m_i, m_j) \geq \tilde{W}_i(m_i, m_j, t_i)$ and $w_i = 1$ otherwise.

The choices of w_i are best responses by construction. The condition of the proposition implies $V_2(0,0) \geq \mathrm{RV}_2(\bar{t}_2, 0)$, so all types of both players choose $w_i = 0$ following $(m_1, m_2) = (0,0)$. The updated beliefs following $m_j = 0$ are in accordance with Bayes' rule, and all other beliefs are unrestricted by perfect Bayesian equilibrium. Finally, it is unprofitable for any type to deviate to $m_i > 0$, as doing so yields an expected utility of

$$\begin{aligned}
\max\left\{V_i(m_i, 0), \tilde{W}_i(m_i, 0, t_i)\right\} &\leq \max\left\{V_i(0,0), \mathrm{RV}_i(t_i, 0)\right\} \\
&\leq \max\left\{V_i(0,0), \mathrm{RV}_i(\bar{t}_i, 0)\right\} \\
&= V_i(0,0). \qquad \square
\end{aligned}$$

A.2 Proof of Proposition 2

We begin with a series of lemmas concerning players' reservation values and their minimization. The first derives a player's optimal investment if she expects to force a conflict, given investment $m_j > 0$ by the other player.

Lemma 1. *For all* $m_j \in (0, y_j]$, *let*

$$\mathrm{BR}_i(m_j) = \min\left\{ y_i, \frac{\sqrt{\theta_i\theta_j m_j + \theta_j(\theta_j - \theta_i)m_j^2} - \theta_j m_j}{\theta_i} \right\}.$$

$\mathrm{BR}_i(m_j)$ *is the unique maximizer of* $\tilde{W}_i(m_i, m_j, t_i)$ *for all types of player* i*; i.e., for all* $t_i \in T_i$ *and* $m_i \in M_i \setminus \mathrm{BR}_i(m_j)$,

$$\tilde{W}_i(\mathrm{BR}_i(m_j), m_j, t_i) > \tilde{W}_i(m_i, m_j, t_i).$$

Proof. Take any $m_i \in [0, y_i]$, $m_j \in (0, y_j]$, and $t_i \in T_i$. Notice that

$$\begin{aligned} \frac{\partial \tilde{W}_i(m_i, m_j, t_i)}{\partial m_i} &= \frac{\partial p_i(m_i, m_j)}{\partial m_i}\left[1 - m_i - m_j\right] - p_i(m_i, m_j) \\ &= \frac{\theta_i\theta_j m_j}{(\theta_i m_i + \theta_j m_j)^2}\left[1 - m_i - m_j\right] - \frac{\theta_i m_i}{\theta_i m_i + \theta_j m_j}. \end{aligned} \tag{3}$$

This expression is strictly decreasing in m_i, which means $\tilde{W}_i$ is strictly concave in m_i. This in turn means that $\tilde{W}_i$ has a unique maximizer with respect to m_i. Because $\partial \tilde{W}_i(0, m_j, t_i)/\partial m_i > 0$, the maximizer is the unique value at which Equation 3 equals zero, or else y_i if the unconstrained maximizer is infeasible. Moreover, because the type t_i does not enter the marginal utility defined by Equation 3, this maximizer is the same for all $t_i \in T_i$.

Let m_i' denote the unconstrained maximizer. By setting Equation 3 to equal zero and rearranging terms, we yield

$$1 - m_i' - m_j = \frac{m_i'\left(\theta_i m_i' + \theta_j m_j\right)}{\theta_j m_j}, \tag{4}$$

which is equivalent to

$$\theta_i(m_i')^2 + 2\theta_j m_j m_i' - \theta_j m_j(1 - m_j) = 0.$$

The quadratic theorem then implies

$$m_i' = \frac{-2\theta_j m_j \pm \sqrt{(2\theta_j m_j)^2 + 4\theta_i\theta_j m_j(1-m_j)}}{2\theta_i}.$$

Because investment cannot be negative, the only valid solution is the positive root, which gives the result. □

We now identify the unique pair of investment levels that form mutual best responses when neither player's wealth constraint binds. Specifically, for each $i = 1, 2$, let

$$m_i^\dagger = \begin{cases} \dfrac{1}{4} & \theta_j = \theta_i, \\ \dfrac{\sqrt{\theta_j}\left(\sqrt{\theta_j} - \sqrt{\theta_i}\right)}{2\left(\theta_j - \theta_i\right)} & \theta_j \neq \theta_i. \end{cases}$$

Corollary 1. $\mathrm{BR}_i(m_j^\dagger) = \min\{m_i^\dagger, y_i\}$.

Proof. Assume $\mathrm{BR}_i(m_j^\dagger) < y_i$. It is immediate from the definition of $m_j^\dagger$ that

$$\theta_i\theta_j m_j^\dagger + \theta_j(\theta_j - \theta_i)(m_j^\dagger)^2 = \frac{\theta_i\theta_j}{4}. \tag{5}$$

In case $\theta_i = \theta_j$, it is then immediate that $\mathrm{BR}_i(m_j^\dagger) = 1/4 = m_i^\dagger$. Otherwise, we have

$$\mathrm{BR}_i(m_j^\dagger) = \frac{1}{\theta_i}\left[\frac{\sqrt{\theta_i\theta_j}}{2} - \frac{\theta_j(\theta_i - \sqrt{\theta_i\theta_j})}{2(\theta_i - \theta_j)}\right] = \frac{\sqrt{\theta_j}(\sqrt{\theta_i} - \sqrt{\theta_j})}{2(\theta_i - \theta_j)} = m_i^\dagger,$$

as claimed. □

The argument in the main text implies that a necessary condition for a peaceful equilibrium with investment levels (m_i, m_j) is

$$\mathrm{RV}_i(\bar{t}_i, m_j) + m_j + \mathrm{RV}_j(\bar{t}_j, m_i) + m_i \leq 1.$$

We define the functions ψ_i and ψ_j such that the above condition is equivalent to

$$\psi_i(m_i) + \psi_j(m_j) \leq 1 + \bar{c}_i + \bar{c}_j - \bar{t}_i - \bar{t}_j.$$

Specifically, let $\psi_i : M_i \to \mathbb{R}$ be defined by

$$\psi_i(m_i) = \begin{cases} 1 & m_i = 0, \\ p_j(\mathrm{BR}_j(m_i), m_i)\,[1 - \mathrm{BR}_j(m_i) - m_i] + m_i & m_i > 0. \end{cases}$$

Notice that ψ_i is continuous, as $\lim_{m_i \to 0^+} \psi_i(m_i) = 1$.

To find the widest conditions under which there exist peaceful equilibria, we are concerned with the minimization of ψ_i. The following result ensures that ψ_i has a unique minimizer.

Lemma 2. *ψ_i is strictly convex.*

Proof. First consider $m_i \in (0,1)$ such that $\mathrm{BR}_j(m_i) < y_j$. Let $g(m_i) = \theta_i\theta_j m_i + \theta_i(\theta_i - \theta_j)m_i^2$, so that

$$\mathrm{BR}_j(m_i) = \frac{\sqrt{g(m_i)} - \theta_i m_i}{\theta_j}$$

by Lemma 1. This implies

$$\begin{aligned} \psi_i(m_i) &= \frac{\theta_j\,\mathrm{BR}_j(m_i)}{\theta_i m_i + \theta_j\,\mathrm{BR}_j(m_i)}\,[1 - m_i - \mathrm{BR}_j(m_i)] + m_i \\ &= \frac{\sqrt{g(m_i)} - \theta_i m_i}{\sqrt{g(m_i)}}\left[\frac{\theta_j(1-m_i) - (\sqrt{g(m_i)} - \theta_i m_i)}{\theta_j} + m_i\right] \\ &= 1 + \frac{2}{\theta_j}\left[\theta_i m_i - \sqrt{g(m_i)}\right]. \end{aligned}$$

This in turn gives

$$\psi_i'(m_i) = \frac{2}{\theta_j}\left[\theta_i - \frac{g'(m_i)}{2\sqrt{g(m_i)}}\right]$$

and thus

$$\psi_i''(m_i) = \frac{g'(m_i)^2 - 2g(m_i)g''(m_i)}{2\theta_j g(m_i)\sqrt{g(m_i)}}.$$

The denominator of this expression is positive, so its sign equals that of

$$g'(m_i)^2 - 2g(m_i)g''(m_i) = (\theta_i\theta_j)^2 > 0.$$

Now consider $m_i \in (0,1)$ such that $\mathrm{BR}_j(m_i') = y_j$ for all m_i' in a neighborhood of m_i. Here we have

$$\psi_i(m_i) = \frac{\theta_j y_j}{\theta_i m_i + \theta_j y_j}[1 - m_i - y_j] + m_i,$$

yielding the derivative

$$\psi_i'(m_i) = \frac{-\theta_i\theta_j y_j}{(\theta_i m_i + \theta_j y_j)^2}[1 - m_i - y_j] - \frac{\theta_j y_j}{\theta_i m_i + \theta_j y_j} + 1.$$

It is clear that this expression is strictly increasing in m_i, so $\psi_i''(m_i) > 0$. $\square$

This result allows us to characterize the minimizer of $\psi_i(m_i) + \psi_j(m_j)$ both when the wealth constraint does not bind and when it does.

Lemma 3. *If $y_i \geq m_i^\dagger$ and $y_j \geq m_j^\dagger$, then*

$$\min_{(m_i,m_j)\in M_i\times M_j}\left\{\psi_i(m_i) + \psi_j(m_j)\right\} = \psi_i(m_i^\dagger) + \psi_j(m_j^\dagger) = 1.$$

Proof. Assume $y_i \geq m_i^\dagger$ and $y_j \geq m_j^\dagger$. First we show that $m_i^\dagger$ is the unconstrained minimizer of ψ_i. Lemma 2 implies that ψ_i has a unique minimizer and that $\psi_i'(m_i) = 0$ is a sufficient condition for m_i to be the minimizer. Let $g(m_i)$ be defined as in the proof of Lemma 2, so we have

$$\psi_i'(m_i^\dagger) = \frac{2}{\theta_j}\left[\theta_i - \frac{\theta_i\theta_j + 2\theta_i(\theta_i - \theta_j)m_i^\dagger}{\sqrt{\theta_i\theta_j}}\right].$$

In case $\theta_i = \theta_j$, then clearly $\psi_i'(m_i^\dagger) = 0$, as required. Otherwise, if $\theta_i \neq \theta_j$, we have

$$\theta_i\theta_j + 2\theta_i(\theta_i - \theta_j)m_i^\dagger = \theta_i\theta_j + \theta_i\sqrt{\theta_j}(\sqrt{\theta_i} - \sqrt{\theta_j}) = \theta_i\sqrt{\theta_i\theta_j},$$

so $\psi_i'(m_i^\dagger) = 0$, again as required.

To prove that the minimized value is 1, observe that

$$\begin{aligned}
\theta_i^2 m_i^\dagger + \theta_j^2 m_j^\dagger &= \begin{cases} \dfrac{\theta_i^2 + \theta_j^2}{4} & \theta_i = \theta_j, \\ \dfrac{(\theta_i^2\sqrt{\theta_j} + \theta_j^2\sqrt{\theta_i})(\sqrt{\theta_j} - \sqrt{\theta_i})}{2(\theta_j - \theta_i)} & \theta_i \neq \theta_j \end{cases} \\
&= \frac{(\theta_i + \theta_j)\sqrt{\theta_i\theta_j} - \theta_i\theta_j}{2}
\end{aligned}$$

and therefore

$$\begin{aligned}
\psi_i(m_i^\dagger) + \psi_j(m_j^\dagger) &= 2 + 2\left[\frac{\theta_i m_i^\dagger - \sqrt{g(m_i^\dagger)}}{\theta_j} + \frac{\theta_j m_j^\dagger - \sqrt{g(m_j^\dagger)}}{\theta_i}\right] \\
&= 2 + 2\left[\frac{\theta_i^2 m_i^\dagger + \theta_j^2 m_j^\dagger - (\theta_i + \theta_j)\sqrt{\theta_i\theta_j}/2}{\theta_i\theta_j}\right] \\
&= 1,
\end{aligned}$$

as claimed. □

The unconstrained solution is infeasible if either player's wealth is too low. In this case, define

$$m_i^\ddagger = \mathrm{BR}_i(y_j) = \frac{\sqrt{\theta_i\theta_j y_j + \theta_j(\theta_j - \theta_i)y_j^2} - \theta_j y_j}{\theta_i}.$$

Lemma 4. *If $y_j < m_j^\dagger$, then $y_i > m_i^\dagger$ and*

$$\min_{(m_i, m_j) \in M_i \times M_j} \Big\{\psi(m_i) + \psi(m_j)\Big\} = \psi_i(m_i^\ddagger) + \psi_j(y_j) = 1.$$

Proof. Assume $y_j < m_j^\dagger$. Because $m_k^\dagger < 1/2$ for each $k = 1, 2$ and $y_i + y_j = 1$, $y_j < m_j^\dagger$ implies $y_i > m_i^\dagger$. As ψ_j is strictly convex, per Lemma 2, its constrained minimizer on M_j is y_j. From there we must show that $m_i^\ddagger$ minimizes ψ_i. It cannot be minimized at a value at which player j plays her unconstrained best response, as the proof of Lemma 3 implies that

$m_i^\dagger$ is the only such minimizer. Solving the minimization condition $\psi_i'(m_i) = 0$ for m_i such that $\mathrm{BR}_j(m_i) = y_j$ in a neighborhood of m_i yields $m_i = m_i^\ddagger$. Finally, we have

$$\begin{aligned}
\psi_i(m_i^\ddagger) &= \frac{\theta_j y_j}{\sqrt{\theta_i\theta_j y_j + \theta_j(\theta_j - \theta_i)y_j^2}} \\
&\quad \times \left[\frac{\theta_i - \theta_i y_j + \theta_j y_j - \sqrt{\theta_i\theta_j y_j + \theta_j(\theta_j - \theta_i)y_j^2}}{\theta_i}\right] \\
&= \frac{2}{\theta_i}\left[\sqrt{\theta_i\theta_j y_j + \theta_j(\theta_j - \theta_i)y_j^2} - \theta_j y_j\right], \\
\psi_j(y_j) &= 1 + \frac{2}{\theta_i}\left[\theta_j y_j - \sqrt{\theta_i\theta_j y_j + \theta_j(\theta_j - \theta_i)y_j^2}\right],
\end{aligned}$$

so $\psi_i(m_i^\ddagger) + \psi_j(y_j) = 1$, as claimed. □

We can now prove the proposition.

Proposition 2. *There are cost thresholds for a monopoly of violence, $\bar{\psi}_1$ and $\bar{\psi}_2$, such that:*

(a) $0 < \bar{\psi}_1 \leq 1/2$ *and* $\bar{\psi}_2 = 1 - \bar{\psi}_1 \geq 1/2$.

(b) If $\bar{\psi}_i + \bar{t}_i + \bar{t}_j \leq \bar{c}_i + \bar{c}_j < 1 + \bar{t}_i + \bar{t}_j$, then there is a peaceful equilibrium with a monopoly of violence by player i if and only if initial wealth is not too skewed in favor of player j.

(c) If $\bar{t}_i + \bar{t}_j < \bar{c}_i + \bar{c}_j < \bar{\psi}_i + \bar{t}_i + \bar{t}_j$, then there is a peaceful equilibrium with a monopoly of violence by player i if and only if initial wealth is skewed far enough in favor of player i.

Proof. We define each cost threshold as $\bar{\psi}_i = 2m_i^\dagger$, from which part (a) follows.

To prove that the stated conditions are necessary, recall that a monopoly of violence by player i requires that

$$\psi_i(m_i) + \psi_j(0) \leq 1 + \bar{c}_i + \bar{c}_j - \bar{t}_i - \bar{t}_j,$$

or, equivalently,

$$\bar{c}_i + \bar{c}_j \geq \psi_i(m_i) + \bar{t}_i + \bar{t}_j \tag{6}$$

for some $m_i \in M_i$. If the condition of part (b) holds, then from Lemma 3 the necessary condition is equivalent to

$$\psi_i(\min\{m_i^\dagger, y_i\}) \leq \bar{c}_i + \bar{c}_j - \bar{t}_i - \bar{t}_j,$$

which holds if and only if y_i is great enough (and thus y_j small enough). If the condition of part (c) holds, then Lemma 3 implies that the necessary condition cannot be met at an investment level at which neither player's wealth constraint binds. In this case, then, Lemma 4 implies that the necessary condition is equivalent to

$$\psi_i(m_i^\ddagger) \leq \bar{c}_i + \bar{c}_j - \bar{t}_i - \bar{t}_j,$$

which holds if and only if y_j is small enough (and thus y_i great enough).

To prove sufficiency, we construct the claimed equilibrium. Assume the necessary condition holds, and take any $\tilde{m}_i \in M_i$ such that $\bar{c}_i + \bar{c}_j \geq \psi_i(\tilde{m}_i) + \bar{t}_i + \bar{t}_j$. Define the redistribution scheme V_i as follows:

- $V_i(m_i, 0) = \mathrm{RV}_i(\bar{t}_i, 0) - \max\{m_i - \tilde{m}_i, 0\}$.
- For all $m_j > 0$, $V_i(m_i, m_j) = V_i(\tilde{m}_i, 0)$.

We then claim the following assessment constitutes a monopoly of violence equilibrium:

- Every type of player i chooses $m_i = \tilde{m}_i$.
- Every type of player j chooses $m_j = 0$.
- After observing the other player's investment choice, each player's updated belief about the other's type equals her prior.
- After investment choices (m_i, m_j), type t_k of player k chooses $w_k = 0$ if $V_k(m_i, m_j) \geq \tilde{W}_k(m_i, m_j, t_k)$ and $w_k = 1$ otherwise.

The proof that this is an equilibrium is analogous to the proof of Proposition 1. As in that proof, the choices of w_k are best responses by construction, and the beliefs are updated in accordance with Bayes' rule whenever possible. Our conditions on $\tilde{m}_i$ imply that $V_j(0, \tilde{m}_i) = 1 - \tilde{m}_i - \mathrm{RV}_i(\bar{t}_i, 0) \geq \mathrm{RV}_j(\bar{t}_j, \tilde{m}_i)$, so there is peace along the path of play. Finally, the redistribution scheme is designed such that neither player has a unilateral incentive to deviate from the prescribed investment choice. □

A.3 Proof of Proposition 3

Proposition 3. *In a monopoly of violence, the monopolist prefers more coercive investment than the socially efficient level.*

Proof. Assume there is an equilibrium with a monopoly of violence by player i, per the conditions of Proposition 2. The monopolist's greatest feasible payoff from such an equilibrium is

$$V_i(m_i, 0) = 1 - m_i - \mathrm{RV}_j(\bar{t}_j, m_i) = 1 - \psi_i(m_i).$$

The best equilibrium for the monopolist is therefore the one that minimizes $\psi_i(m_i)$, subject to the constraint of Equation 6. By contrast, the most socially efficient monopoly of violence is the lowest value of m_i at which Equation 6 holds with equality. Unless there is only one investment level satisfying the constraint, Lemma 2 implies that the constrained minimizer of $\psi_i(m_i)$ is greater than the socially efficient level. □

A.4 Proof of Proposition 4

Proposition 4. *There is a peaceful equilibrium with a balance of power if and only if* $\bar{c}_1 + \bar{c}_2 \geq \bar{t}_1 + \bar{t}_2$.

Proof. By arguments similar to those in the proof of Proposition 2, the necessary condition for a peaceful balance of power is the existence of $m_i \in M_i$ and $m_j \in M_j$ such that

$$\bar{c}_i + \bar{c}_j \geq \psi_i(m_i) + \psi_j(m_j) - 1 + \bar{t}_i + \bar{t}_j. \tag{7}$$

It follows from Lemma 3 and Lemma 4 that such values exist if and only if the condition of the proposition holds. One can then construct an equilibrium analogous to the one constructed in the proof of Proposition 2 to prove sufficiency. □

A.5 Proof of Proposition 5

Proposition 5. *If there is a peaceful equilibrium with a monopoly of violence, then there is a peaceful equilibrium with a balance of power with strictly less total coercive investment.*

Proof. The efficient equilibrium solves the constrained maximization problem

$$\begin{aligned} \max_{m_i, m_j} \quad & 1 - m_i - m_j \\ \text{s.t.} \quad & \psi_i(m_i) + \psi_j(m_j) \leq 1 - \bar{t}_i - \bar{t}_j + \bar{c}_i + \bar{c}_j, \\ & m_i \geq 0, \\ & m_j \geq 0, \\ & m_i \leq y_i, \\ & m_j \leq y_j. \end{aligned}$$

This is a concave maximization problem with a convex constraint set.[23] Because $\lim_{m_j \to 0^+} \psi_j'(m_j) = -\infty$, this cannot be solved with any (m_i, m_j) such that $m_i > 0$ and $m_j = 0$. □

[23]The convexity of the first constraint follows from Lemma 2.

A.6 Proof of Proposition 6

Proposition 6. *The best peaceful equilibrium for player i is a monopoly of violence by i with more coercive investment than is socially efficient, if such an equilibrium exists.*

Proof. As in the proof of Proposition 3, player i's objective is to maximize her feasible payoff,

$$V_i(m_i, m_j) = 1 - m_i - m_j - \psi_i(m_i),$$

subject to the equilibrium constraint of Equation 7. Clearly the optimal solution entails $m_j = 0$ if such an investment profile can Equation 7; i.e., if Equation 6 can be satisfied. From there the result follows from Proposition 3. $\square$

A.7 Equilibria with Conflict

The following proposition formally proves the existence and asymptotic efficiency of the type of equilibrium described in the text.

Proposition 7. *If $\bar{c} \geq 1/[2(1+\pi)^2]$, then there is a conditional monopoly of violence in the symmetric two-type model. As the prior probability of a high type π approaches zero, the inefficiency of this equilibrium approaches zero.*

Proof. Assume the condition of the proposition holds. Let $m_1^* = \pi/(1+\pi)^2$ and $m_2^* = \pi^2/(1+\pi)^2$. Define the redistribution scheme as follows:

- $V_2(0, m_2) = \max\{1 - m_2 - \mathrm{RV}_1(0, m_2^*), 0\}$ for all $m_2 \geq 0$.
- $V_2(m_1, m_2) = (1 - m_1 - m_2)/2$ for all $m_1 > 0$ and $m_2 \geq 0$.

We claim that the following strategy profile constitutes a conditional monopoly of violence equilibrium:

- Player 1 invests $m_1 = 0$ if her type is $t_1 = 0$ and invests $m_1 = m_1^*$ if her type is $t_1 = \bar{c}$.

- Player 2 invests $m_2 = m_2^*$.
- After observing $m_1 = 0$, player 2 infers for certain that $t_1 = 0$. After observing any $m_1 > 0$, player 2 infers for certain that $t_1 = \bar{c}$.
- After the mobilization choice $(0, m_2^*)$, all types of both players participate in the institution.

  After any mobilization choice $(m_1, m_2) \neq (0, m_2^*)$, all types of both players choose to opt for conflict.

The beliefs are consistent with the application of Bayes' rule wherever possible. In the cases where both sides choose conflict, this is trivially an equilibrium—conflict is unilateral, so neither player's decision is pivotal, making $w_i = 1$ trivially a best response for each. In the case where conflict does not occur, we have $V_1(0, m_2^*) = \mathrm{RV}_1(0, m_2^*) > 0 \geq \tilde{W}_1(0, m_2, t_1)$ for all t_1, so player 1's strategy is a best response. In addition Lemma 1 gives

$$\mathrm{RV}_1(0, m_2^*) = \frac{\mathrm{BR}_1(m_2^*)^2}{m_2^*} - \bar{c} = \frac{1}{(1+\pi)^2} - \bar{c}.$$

It is then immediate from $\bar{c} > 0$ that $\mathrm{RV}_1(0, m_2^*) < 1 - m_2^*$, so we have $V_2(0, m_2^*) = 1 - m_2^* - \mathrm{RV}_1(0, m_2^*)$. In order for player 2's participation in the institution here to be a best response, we must have $V_2(0, m_2^*) \geq 1 - m_2^* - \bar{c}$; this is equivalent to $\mathrm{RV}_1(0, m_2^*) \leq \bar{c}$, which in turn is equivalent to the condition of the proposition.

The last step to confirm that this is an equilibrium is to confirm that each type's investment strategy is optimal. There is clearly no profitable deviation for type $t_1 = 0$ of player 1, as this type receives its reservation value along the equilibrium path, and any deviation would result in conflict. There is also no profitable deviation for type $t_1 = \bar{c}$ of player 1. Its strategy is optimal in case of conflict, as $m_1^* = \mathrm{BR}_1(m_2^*)$, and deviating to $m_1 = 0$ would result in a payoff of $\mathrm{RV}_1(0, m_2^*) < \mathrm{RV}_1(\bar{c}, m_2^*)$. Finally, we must confirm that player 2's strategy is optimal. In expectation the most it can receive from deviating and forcing a conflict is

$$\sup_{m_2} \left\{ (1-\pi) \left[1 - m_2 - \bar{c} \right] + \pi \left[p_2(m_1^*, m_2)(1 - m_1^* - m_2) - 2\bar{c} \right] \right\},$$

which is maximized at $m_2 = m_2^*$. Therefore, player 2's investment strategy is optimal as well.

It is evident that $m_1^* \to 0$ and $m_2^* \to 0$ as $\pi \to 0$. The probability of costly conflict also goes to zero as $\pi \to 0$, as conflict occurs on the path of play only if $t_1 = \bar{c}$. This proves the second claim of the proposition. □

A.8 Exit

We first state and prove a lemma about the necessary condition for an equilibrium with a monopoly of violence in the game with exit.

Lemma 5. *In the game with exit, there is an equilibrium with a monopoly of violence by player i with mobilization level $m_i \geq 0$ only if*

$$\max \left\{ 1 - \bar{c}_i + \bar{t}_i, \alpha y_i \right\} + \max \left\{ \mathrm{RV}_j(\bar{t}_j, m_i), \alpha y_j \right\} \leq 1 - m_i.$$

Proof. The strongest type of the monopolist's conflict constraint is $V_i(m_i, 0) \geq 1 - \bar{c}_i + \bar{t}_i$, and her exit constraint is $V_i(m_i, 0) \geq \alpha y_i$. Similarly, the strongest type of the subject's conflict constraint is $V_j(m_i, 0) \geq \mathrm{RV}_j(\bar{t}_j, m_i)$, and her exit constraint is $V_j(m_i, 0) \geq \alpha y_j$. As the redistributive offer must satisfy $V_i(m_i, 0) + V_j(m_i, 0) = 1 - m_i$, this concludes the proof. □

This is the foundation for the efficiency result noted in the text.

Proposition 8. *In the game with exit, if there is an equilibrium with a monopoly of violence by player i and $\alpha y_i > 1 - \bar{c}_i + \bar{t}_i$, then this equilibrium is less efficient than would be possible in the baseline game.*

Proof. Let m_i be the investment level in the most efficient monopoly of violence in the baseline game; i.e., let m_i be the minimal solution to

$$1 - \bar{c}_i + \bar{t}_i + \mathrm{RV}_j(\bar{t}_j, m_i) - \bar{c}_j + \bar{t}_j = 1 - m_i.$$

Therefore, if $\alpha y_i > 1 - \bar{c}_i + \bar{t}_i$, then the condition of Lemma 5 cannot hold at m_i. $\square$