GROUPS IN CONFLICT:

Size Matters, But Not In The Way You Think[1]

LAURA MAYORAL
Instituto de Análisis Económico (CSIC) and Barcelona GSE

DEBRAJ RAY
New York University and University of Warwick

January 2016, revised October 2016

This paper studies costly conflict over private and public goods. Oil is an example of the former, political power an example of the latter. Groups involved in conflict are likely to be small when the prize is private, and large when the prize is public. We examine these implications empirically by constructing a global dataset at the ethnic group level and studying conflict along ethnic lines. Our theoretical predictions find significant confirmation in an empirical setting.

1. INTRODUCTION

We study a model of social conflict in which there are multiple potential threats to peace. There are several potential groups, demarcated by one or more characteristics —- economic, ethnic, occupational or geographic. From these, a group might emerge to challenge the existing state of affairs. We address three issues:

1. Whether large groups or small groups are more likely to be involved in conflict against the State;

2. When conflict might occur, even if it is costly and inefficient, and even though there is flexibility in the distribution of peacetime allocations; and

3. Whether our predictions regarding group size and conflict are supported by the data.

Which groups are likely to be involved in conflict? This is, of course, a question that cannot be answered in full generality, as the questions of identity and cohesion of various potential groupings are deep issues that can only be resolved through specific econometric and ethnographic research. But there is one aspect of a group that commands special attention, and that can be examined both theoretically and empirically: group *size*. Are large groups or small groups more likely to initiate conflict, or resist what are perceived to be the unfair incursions of the State?

The literature offers both answers. We are all aware of the "tyranny of the majority" (see, e.g. Tocqueville 1835), in which a larger group can impose its will on society even on issues that a relative minority might feel very strongly about. The tyranny expresses itself most clearly in a voting context, for after all, voting is an expression of ordinal preferences, and not the intensity

[1]We are grateful to Oeindrila Dube, William Easterly, Joan Esteban and Sahar Parsa for helpful comments on an earlier draft. Ray thanks the National Science Foundation for research support under grant number SES-1261560.

of those preferences. But it is certainly not limited to voting. The suppression of minorities via extra-democratic channels, including coercive and violent means, is extremely common.

But there is a contrasting view which argues that small groups may be more involved than large groups in lobbying or conflict (see Pareto 1927 and Olson 1965). For more on these matters, see, e.g., Chamberlin (1974), McGuire (1974), Marwell and Oliver (1993), Oliver and Marwell (1988), Sandler (1992), Taylor (1987) and Esteban and Ray (2001a). This literature studies intensity of conflict displayed across small and large groups, *assuming that there is conflict to begin with*. In contrast, we ask whether a small or large group is willing to enter a conflict, or to resist a perceived act of aggression. This is a more subtle issue — after all, it is generally the case that large groups continue to have better chances of winning the conflict. But the group with the better chances is not necessarily the one to get involved. Rather, the entry into conflict depends on the expected payoff to a group, *relative to its received allocation otherwise*.

For concreteness, suppose that an ethnic group in a country has oil reserves located in its homeland. Suppose that these are distributed equally across the entire country. Or suppose that the homeland itself is settled by other ethnicities in the country. If, as a result of redistribution or settlement, the resources are divided uniformly across the country, the revenue share of our ethnic group, with population share n, will be just n. If that group is involved in a war of secession, taking the oil or land with it in the event of victory, its chances of winning will be some function $p(n)$ (to be computed). The question is not whether $p(n)$ is small or large, but how large it is *relative* to n (and net of any costs of conflict). That will determine the decision to get involved in a conflict — either to initiate, or depending on the context at hand, to resist.

Our exercise has a sharp implication: under the assumption that the peacetime allocation is uniform (that is, contestable resources are equally allocated), conflict is more likely to be associated with small groups when the prize in question is private, but more likely to be associated with large groups when the prize is public. See Propositions 1 and 2. This is the central prediction that we take to the data.

The analysis above can be developed further.[2] The central difficulty to peace is not the presence of a single threat from a small or large group which can be appeased in Coaseian fashion, but the existence of *several* conflictual divisions of society, each based on a different marker, such as class, geography, religion, or ethnicity. Because conflict is inefficient, society can arrange — for every potential conflict — a set of Coaseian transfers that Pareto-dominate the expected payoffs *under that conflict*. But it may be unable to find an arrangement that *simultaneously* prevents *all* such threats to peace.[3] For instance, a society can set up institutions that can adequately deal with

[2]While we do not subject these additional results to empirical scrutiny, we hope that finer datasets available in the future will permit such an analysis.

[3]There is a large literature on the persistence of costly conflict. This literature invokes the usual suspects; principally, adverse selection and moral hazard. Fearon (1995), Esteban and Ray (2001b), Baliga and Sjöstrom (2004), Bester and Warneryd (2006) and Sánchez-Pagés (2009) all employ some version of a hidden-type model to obtain their results. For instance, both parties might feel they have the better chance of prevailing in the conflict. Fearon (1995), Garfinkel and Skaperdas (2000), Powell (2004, 2006), Slantchev (2003) and Jackson and Morelli (2007) use different variants of a moral hazard framework. For instance, an allocation that Pareto-improves upon the conflict outcome will generally require transfers to implement, but there is no guarantee that those transfers will be actually implemented ex post. In contrast, we focus on the multiplicity of threats in a complete-information setting.

the question of class conflict, only to be confronted by threats from a religious or geographical subgroup.[4]

It is, of course, possible that different markers all delineate essentially the same division of people: e.g., “poor” and “rich” might generate the same division as “North” and “South”. In that case our argument fails. In contrast, the argument is strongest when the different markers generate “orthogonal” divisions of society.[5] Then a system of transfers set up to deal with one sort of division may be useless when confronted with another. We provide conditions under which conflict is inevitable, in the sense that there is no allocation — uniform or not — that can simultaneously handle all competing demands. See Propositions 4 and 5 in Section 4.

In the second part of the paper, we empirically test our predictions on group size, the nature of the prize, and conflict. To this end, we focus on groups that are defined along ethnic lines. Ethnic conflict is a natural choice for the study, as groups demarcated by ethnicity account for between 50–75% of internal conflicts since 1945 (Fearon and Laitin, 2003; Doyle and Sambanis, 2006). To carry out the analysis, we construct a panel dataset at the ethnic group level with global coverage. The dataset contains information for 145 countries and 1475 ethnic groups spanning the years 1960 to 2006.

The data is replete with examples of both public- and private-goods conflict; often mixtures of the two to be sure. The typical ethnic conflict could involve a struggle for political power or control (as in Burundi, Bosnia, Liberia, or Zimbabwe), but it can involve secessionist struggles by groups seeking to control their own land or resources (Tamils in Sri Lanka, the Casamance in Senegal, Chechnya, or various separatist movements in India). Land and oil are often central among these resources (e.g., the Ijaw conflict in Nigeria, the Darfur conflict, or the Second Civil War in the Sudan). Our empirical strategy, which we discuss in more detail later, is to allow for possible mixtures and yet tease out private and public components of the conflict.

To obtain a proxy for private payoffs, we consider rents that are easily appropriable. Because appropriability is closely connected to the presence of resources, we approximate the degree of “privateness” in the prize by asking if the homeland of the ethnic group is rich in natural resources. In our baseline specification we use oil abundance in the homeland as a proxy of “privateness,” but we also consider alternative measures based on mineral and land abundance, again at the ethnic group level.

We approach the notion of “publicness” in two ways. The first is a specific measure of pre-sample autocracy constructed by Polity IV, which is a country-level index based on the degree of power afforded to those who run the country. Our underlying idea is that if the State is classified as autocratic to begin with, there will more to gain for a group by seizing power. Or it may be that the disaffected who seize power simply want to get rid of the government and start a transition to

[4]For instance, in India, several groups have challenged the center in conflictual situations: a casual list would include fundamentalists (both Hindus and Muslims), revolutionary groups based on class (such as the Naxalites), high-caste groups, the scheduled castes, geographical areas such as the North East States or the Punjab, agricultural labor, farmer groups, trade unions, industrial lobbies, and so on.

[5]In this sense, our first question follows the lead of Esteban and Ray (2008), which considers only two markers — class and ethnicity — and assume these to be orthogonal. However, that paper does not consider the question of whether small or large groups initiate conflict.

democracy. Our second approach is to simply treat publicness as a "residual" after the influence of our measure of private payoffs is fully netted out.

Our results appear to firmly support the predictions of the theory: smaller ethnic groups are more likely to be involved in conflict (using both conflict *onset* and conflict *incidence* as measures) when oil, minerals or land are abundant for the group. At the same time, using the specific measure of publicness just described, larger ethnic groups are more likely to participate in conflict when the valuation of the public payoff is high. Moreover, once the private prize and its interaction with group size have been accounted for, the coefficient on group size turns positive and significant (not surprisingly, it is insignificant if entered on its own). That is, if publicness is viewed along the line of the second approach outlined above, there is additional support for the positive association of group size and conflict with public payoffs. Our conclusions are robust to a large number of robustness checks that include the consideration of alternative conflict variables, estimation strategies and ways of proxying for the prizes at stake, both private and public. At various points in the paper, we extensively discuss our baseline empirical specification, and its interpretation.

Of course, it is well known in the empirical literature that the presence of natural resources — particularly oil — is correlated with conflict; see, for example, Le Billon (2001), Fearon (2005), Lujala (2010) and Dube and Vargas (2013). Morelli and Rohner (2015) show, additionally, that the *concentration* of those natural resources in ethnic homelands is related to conflict. As in the Morelli-Rohner paper, our empirical study is set at the ethnic group level. Our focus, however, is on the *interaction* between group size and the homeland resource variable. In addition, as already described, we are equally interested in the public payoff variable and its interaction with group size. To our knowledge, neither interaction has been explored empirically in the literature. Together, they reconcile the Tyranny of the Majority with the Pareto-Olson thesis.

The broader empirical question that motivates us is one of the salience of certain group characteristics in conflict. We would like to know just which groups are vulnerable to — or seek out — conflict in different economic situations. This permits policy-makers to be sensitive to the impact on such groups when designing policy. To be sure, there is no general recipe for conducting this analysis, which must often be society- or country-specific.[6] But one factor that could transcend such specificities is group size. Additionally, size has been a central feature in other models of political economy (the Pareto-Olson argument being a leading example). Hence our interest.

In what follows, Section 2 introduces a baseline model of peace and conflict. Sections 3.1 and 3.2 analyze the relation between group size and conflict when conflict is over private and public goods respectively, and the proposed allocation without conflict is uniform. Section 4 studies conditions for "active conflict" no matter what the allocation is. Section 5 introduces the data employed in the empirical analysis and Section 6 presents our baseline empirical results. Section 7 presents a number of robustness checks. Section 8 concludes.

[6]For instance, Mitra and Ray (2014) make the connection between economic forces and conflict across specific religious groups in India.

2. A BASELINE MODEL OF CONFLICT

2.1. **Allocations.** A society is made up of a unit mass of individuals. Denote by v the total "appropriable resources" of society. The value may be transferable to different degrees; we will return to this point below. For now let V denote the space of all transferable allocations of the form $\mathbf{x} = \{x(i)\}$, where i indexes a person. That is, $V = \{\mathbf{x} | \int x(i) \leq v\}$. An allocation is *uniform* if every individual receives the same amount under it.

2.2. **Conflict.** The value v may represent material or economic resources such as oil from a particular geographical location within the society, or the payoff to acquiring political or cultural power. (There may be other non-appropriable human or physical resources which we normalize to zero for everyone.) Assume that there is a subgroup, demarcated by ethnicity, geography, religion or occupation, which seeks to retain — or seize — the proceeds of v entirely for itself, while the State (or society as a whole) seeks to allocate v more widely over the larger community. For instance, in the case of seizable resources, one might think v as the value of oil reserves located within the homeland of an ethnic group. The State wants to distribute those revenues over the entire country, while the ethnic group might feel that this is "their oil." Or, in the case where v represents political power, it is the joint payoff to a group from seizing power.

There are two possibilities: the group can accede to the no-conflict allocation, or its members can engage in costly conflict to retain private resources, or to seize power. In the case of conflict, we suppose that society is partitioned into two subsets, one of size m (pertaining to the group in question) and the remainder of size $\overline{m}$ ($m + \overline{m} = 1$), and that they engage in a bilateral conflict. In short, our group does battle against society as a whole, with the complementary group being a proxy for the incumbent government. We leave open the interpretation of whether our group "initiates" conflict or "defends itself" against what it perceives to be the incursions made by the State. That will depend on the situation at hand. For instance, if there is settlement on the group's territory, conflict may be interpretable as defense against State aggression. If the group is fighting to overthrow the State and seize power, then the group may be viewed as the aggressor. We sidestep these interpretations altogether and simply refer to the two groups as "Rebel" and "State."

Conflict involves — on each side — the expending of effort or resources. The utility cost to an individual from a contribution of r is given by

$$c(r) = (1/\alpha) r^{\alpha}$$

for some $\alpha > 1$. We will presume that if the Rebel seizes the aggregate value v, it can distribute that value among its members in an entirely costless fashion. The same is true of the State.[7] Therefore, it is assumed that a leader on each side extracts these resources from everyone to maximize the per-capita payoff of her coalition.[8] Because the cost of effort provision is strictly

[7] Assume that the State entirely excludes all members of the Rebel in case the Rebel is defeated.

[8] To be sure, this neglects the free-rider problem or the question of intra-group *cohesion*, which is another aspect of small versus large groups worth studying, though we don't do so here. It is easy to write down variants of our model in which individuals unilaterally make resource contributions, provided that they at least partially internalize the payoffs of their fellow group members (see Esteban and Ray 2011).

convex, the leader will ask for equal effort from each individual, and will make transfers if needed to compensate them.

A remark on functional forms: nothing of substance hangs on the specific choice of the cost function. Strict convexity of cost is essential, however.

To map efforts into win probabilities, we use contest success functions (Skaperdas 1996), so the probability that the Rebel will win is given by

$$p = \frac{mr}{R},$$

where r is contribution per person in the Rebel, and $R = mr + \overline{mr}$ is the sum of contributions made by both the groups. (Throughout, we use bars on the corresponding variables for the State.) As in the case of the cost function, this specification too can be substantially generalized.

The Rebel seeks to maximize its per-capita payoff

$$\pi \frac{mr}{R} - c(r),$$

where π is the per-capita value of the prize for it. A similar problem is faced by the State, with a per-capita prize of $\overline{\pi}$.[9] In fact, once conflict has been "declared," the qualitative situation is symmetric for both the groups. A conflict equilibrium is just a Nash equilibrium of this game. Such equilibria are fully described by the first-order conditions

$$\pi m \overline{m} = R^2 \frac{r^{\alpha-1}}{\overline{r}} \tag{1}$$

for the Rebel, and by

$$\overline{\pi} m \overline{m} = R^2 \frac{\overline{r}^{\alpha-1}}{r} \tag{2}$$

for the State. Conditions (1) and (2) yield a simple expression for the provision of individual resources by the group, relative to its rival:

$$\frac{r}{\overline{r}} = \left(\frac{\pi}{\overline{\pi}}\right)^{1/\alpha} \equiv \gamma. \tag{3}$$

We can use these conditions to describe the conflict payoff of each group. For the Rebel, rewrite (1) to observe that

$$r^\alpha = \pi p \overline{p},$$

so that the expected payoff from conflict is given by

$$w \equiv \pi p - c(r) = \pi p - (1/\alpha)\pi p \overline{p} = \pi[kp + (1-k)p^2], \tag{4}$$

where $k \equiv (\alpha - 1)/\alpha$, which lies in $(0, 1)$. Finally, note that

$$p = \frac{mr}{mr + (1-m)\overline{r}} = \frac{m\gamma}{m\gamma + (1-m)}, \tag{5}$$

where γ is defined in (3). Together, (3), (4) and (5) describe a full solution to the Rebel's payoff in conflict equilibrium. A parallel expression holds for the State.

[9] Note that the payoff from defeat is normalized to zero for either group. This will need additional discussion in the case of public goods; see below.

Conflict is a threat to peace, and we seek conditions under which that threat might manifest itself. Say that an allocation $\mathbf{x} \in V$ is *blocked* if the expected payoff to the Rebel under conflict exceeds what it receives under the allocation:

$$\pi[kp + (1-k)p^2] > \int_G x(i).$$

A society is *prone to conflict* if the uniform allocation is blocked. It is *actively conflictual* if every allocation is blocked by *some* Rebel.

For now we focus on proneness. We wish to understand whether small or large Rebels are more likely to be involved in conflict under the imposition of an uniform allocation. To do so, we must compare the payoffs of uniform allocations with those arising from conflict. This is a question that cannot be answered at the current level of generality, for the simple reason that the current model nests both the "tyranny of the majority" as well as the Pareto-Olson argument. The prediction that we generate is therefore subtler than a "one-size-fits-all" answer. We must first link the appropriable surplus v to the prizes π and $\overline{\pi}$. We do so by conducting the exercise in more detail for two leading cases: one in which the prize is a divisible, private good, and the other in which the prize must be used to provide public goods. As we shall see, the answer will be different in each case. It is this leading prediction of the model that we subsequently take to the data.

3. Group Size and Conflict

3.1. **Private Goods.** Little by way of additional interpretation is needed when the entire prize v is a private good; say, oil located on the homeland of the (potential) Rebel. We continue to assume that payoffs are fully transferable, so that in peacetime the prize can be allocated any way we please. In particular, because overall population is normalized to one, the uniform allocation gives precisely v to each individual in society.

Under conflict, then, v represents the total resources at stake. We assume that the winning group seizes the resources entirely. So with a Rebel of size m,

$$\pi = v/m \text{ and } \overline{\pi} = v/(1-m).$$

Using this information in (3), we see that

$$\gamma = \left(\frac{1-m}{m}\right)^{1/\alpha},$$

so that by (5),

$$p = \frac{m^k}{m^k + (1-m)^k}, \tag{6}$$

where $k = (\alpha - 1)/\alpha$.

Notice from (6) that smaller Rebels are disadvantaged in conflict in the sense that they have a lower probability of winning; after all p is increasing in m and $p(1/2) = 1/2$. Nevertheless,

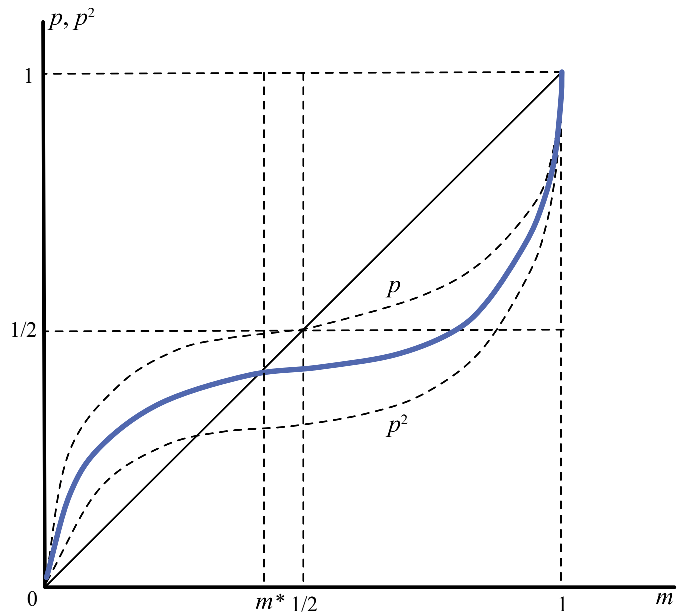


Figure 1. The Threshold m^* in Proposition 1

Proposition 1. *Assume that the budget is private, and the peacetime allocation is uniform. Then there exists* $m^* \in (0, 1/2)$ *such that a Rebel with* $m < m^*$ *will wish to initiate conflict. Society is conflict-prone in the presence of small Rebels.*

The proof that follows may be worth reading as part of the text, as it also provides intuition and tells us how m^* is calculated.

Proof. When the peacetime allocation is uniform, then each individual gets v. On the other hand, using (4), we see that conflict payoff is given by $w = \pi[kp + (1-k)p^2] = v[kp + (1-k)p^2]/m$, so that a Rebel of size m will initiate if

(7) $$kp(m) + (1-k)p(m)^2 > m,$$

where $p(m)$ is given by (6).

The function p has a "reverse-logistic" shape. It starts above the 45^0 line and at the point $n = 1/2$ crosses it and dips below. The derivatives at the two ends are infinite.[10] See Figure 1, which plots p, p^2 and the convex combination $kp + (1-k)p^2$. With this shape in mind, observe that the left-hand side of (7) starts out (for small values of m) higher than the right-hand side and ends up lower (for values of m close to 1). Moreover,

$$kp(m) + (1-k)p(m)^2 < m,$$

[10]To check these claims, note that $\frac{m^k}{m^k+(1-m)^k} \geq n$ if and only if $m \leq 1/2$ (simply cross-multiply and verify this), and that $p'(m) = \frac{km^{k-1}(1-m)^{k-1}}{[m^k+(1-m)^k]^2}$, which is infinite both at $n = 0$ and $n = 1$.

for *any* $m \geq 1/2$.[11] This argument, in conjunction with Figure 1, shows that there is a unique intersection (crossing from above to below) in the interior of $(0, 1/2)$.[12] The proof of the proposition is now complete. ■

Notice that what matters is *not* the level of win probabilities or whether it increases or falls with group size. In fact, it always increases with size. While small Rebels fight more intensely (the per-capita stakes are higher), this does not overturn the fact they have a lower probability of winning than big groups do. Thus small groups engage in conflict not because they have a high chance of winning. (They don't.) Rather, they do so because they have a high chance of winning *relative* to their share from the uniform allocation. That fact is reflected in the reverse-logistic shape of the win probability, derived in the proof of Proposition 1.

We reiterate that we do not interpret this result as a small Rebel deliberately initiating conflict in some "unprovoked fashion." Indeed, in the empirical implementation below, the prize will refer to resources located in the homeland of some ethnic group. The "uniform allocation," in which a State attempts to control these resources in order to redistribute its revenues to the country at large, can be viewed by the group in question as an unwarranted infringement of its rights (to the resource). In that case, the correct interpretation is not one of conflict initiation, but rather resistance.

We end this section with a caveat: our theory only suggests that small groups are associated with conflict, if the initial allocation is uniform. Of course, this tendency is even more pronounced if society has a reason to favor larger groups to begin with, as it will in a democratic (or voting) scenario. But if the initial allocation is chosen to appease the small groups, then it is the larger groups who will have to pay for that appeasement, and *they* will be the initiators (the proverbial dent in the ball must resurface somewhere). But even in this case, the conflict would involve the small groups. We will return to the issue of multiple threats in Section 4. For now, we only note that our empirical findings are consistent all of the above interpretations.

3.2. **Public Goods.** Suppose that the social budget is only used for public goods. For this section, it will be necessary to take account of several potential Rebel groups, each seeking to allocate the budget in the way they see best. (We will nevertheless nest the analysis in the abstract model of Section 2.2.) Specifically, suppose there is a finite collection of groups that partition society, and that there is a *specific* good for each such group (or equivalently, some optimal group-specific *mix* of goods). Take the production function to be as simple as possible: one unit of the budget produces one unit of any of the group-specific goods. Normalize the total budget to have size one.

If all payoffs are transferable, finding the social value v is easy. Choose any group of *maximal* size, say m_1, and devote the budget entirely to the production of that group-specific good. Compensate all other individuals with suitably chosen transfers of money. More concretely, assume that each person derives utility Ψ per unit from her group-specific public good, when she

[11] Suppose this is false for some $1 > m \geq 1/2$. By the properties of p already established, we know that $m \geq 1/2$ implies $m \geq p(m)$, so that $km + (1-k)m^2 \geq m$, but this can never happen when $m < 1$, a contradiction.

[12] More formally, the derivative of $kp(m) + (1-k)p(m)^2$ is strictly smaller than 1 at any intersection, so that there can be only one intersection; we omit the details.

is a member of that group, and 0 otherwise. (None of the arguments that follow rely on such a stark specification, but it helps to fix ideas.) Then $v = \Psi m_1$. This total can be allocated as we please using financial transfers. Under an uniform allocation, for instance, each person receives precisely the value Ψm_1.

3.2.1. *Winners and Losers.* Notice that under any such allocation, there will be groups that can do better if it had costless access to the budget. In fact, it is easy to see that the per-capita worth of a group (conditional on having preferred access) is just Ψ, which exceeds $\Psi m_1 = v$. Of course, rebellion isn't costless; the conflict has to be won first, and conflict is costly.

If, on the other hand, the State wins the conflict, it will generally face a problem of allocation just as society as a whole did. This problem is solved in the same way as it was solved for the society as a whole: the State will produce the public good corresponding to any group of maximal size within it, and allocate the surplus across all *its* subgroups.

We will continue to assume that the losing coalition obtains zero. This assumption is just a normalization when the State loses, but needs more discussion for the case in which the Rebel loses. It is possible that the State puts resources into a group-specific public good that some members of the defeated group benefit from. We can deal with this issue in several ways, but the one that dovetails perfectly with our approach elsewhere in this paper is to ensure that such "enemy beneficiaries" obtain a net payoff of zero by having them suitably taxed by the State. With this in mind, the per-capita worth of the State (conditional on winning) is $\overline{\pi} = \Psi\mu$, where μ is the relative size of the largest non-Rebel group that intersects the State; i.e., the remainder of society apart from the Rebel that lost. That is, $\mu = m'/(1-m)$, where m' is the size of that largest coalition and m, as before, is the size of the erstwhile Rebel. We reiterate that this largest coalition may or may not be fully contained within the State, but that in either case, the State appropriates the full surplus (by requiring transfers or reparations, if necessary).

3.2.2. *Conflict Proneness.* In what follows, consider the (finite) collection of all potential Rebels. Index them in decreasing order of size, so that in particular, m_1 is the size of the largest group and m_2 the size of the second largest group.

Proposition 2. *Assume that the prize is public. Then society is conflict-prone if and only if*

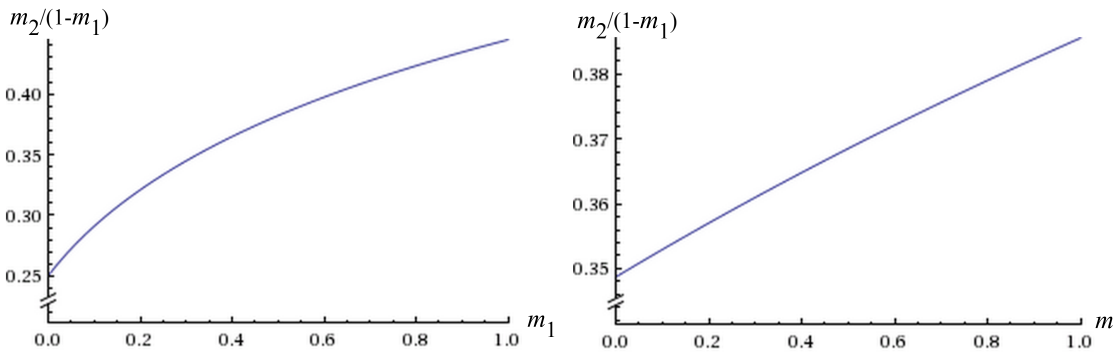
$$m_1 > \frac{1-\gamma_1 k}{(\gamma_1 - 1)^2}, \tag{8}$$

where $\gamma_1 \equiv [(1-m_1)/m_2]^{1/\alpha}$.

Moreover, provided that (8) applies, the largest group prefers conflict to the uniform allocation.

Proof. See Appendix. ■

The condition (8) for conflict-proneness is not just a simple lower bound on m_1. It also depends on the *ratio* of m_1 to m_2. The first point to note is that if there are exactly *two* groups in society, no matter what their relative size, the condition for conflict-proneness is never met. To see this, observe that $\gamma_1 = 1$ in this case, so that the right-hand side of (8) blows up, thereby assuring



(a) $\alpha = 2$ **(b)** $\alpha = 10$

Figure 2. SIZE OF LARGEST GROUP IN POPULATION (m_1) VERSUS SHARE OF SECOND LARGEST GROUP IN REMAINING POPULATION ($m_2/(1-m_1)$)

that the inequality cannot be met. With pure public goods, just two potential Rebels and full transferability, the uniform allocation cannot be blocked. (However, see the remark on limited transferability in Section 3.2.3.)

Matters are different when there are three groups or more. When a Rebel declares conflict against "the rest of society," the State may appear weaker, because the spoils of victory are shared among *its* constituent groups, and therefore do not have the same allure as victory for the initiating Rebel. For instance (and without putting too fine a point on it), the current conflict between the Islamic State and its varied opponents may be tentatively interpreted from this perspective. The more diverse the opposition, the easier it is for a group with common objectives to initiate conflict and bear the costs of that conflict. Figure 2 numerically examines the relationship in (8). The left panel does this for $\alpha = 2$, and the right panel for $\alpha = 10$. The two cases are not too different. In both cases, a potential Rebel of size 50% or more of the population will want to enter into conflict if the second largest group is around 35% or less of the remaining population.

3.2.3. *A Remark on Transferability.* With public goods, we do need to be especially careful about transferability and exactly what it entails. Public goods are not like oil revenues. Think of ethnic or religious representation, or the sharing of political power. The relative price across objects such as these may be very hard to define. So it may be impossible to conceive of "classical" financial transfers as compensation for the loss of power or culture; see, e.g., Kirshner (2000). What price would those who are thus negated accept as compensation?

One way to approach this problem is to allow only for transfers of the *resources* that go into the production of public goods, so that several group-specific goods may need to be produced for allocating payoffs. This is a limited notion of transferability — call it "budget transferability" — and it dilutes aggregate surplus.[13]

[13]To be sure, matters may be more restrictive still. It may be, for instance, that only one public good can be produced at a time — there can only be "one culture." In that case, the payoff frontier with no conflict is just a finite collection of points. We omit this case for reasons of brevity.

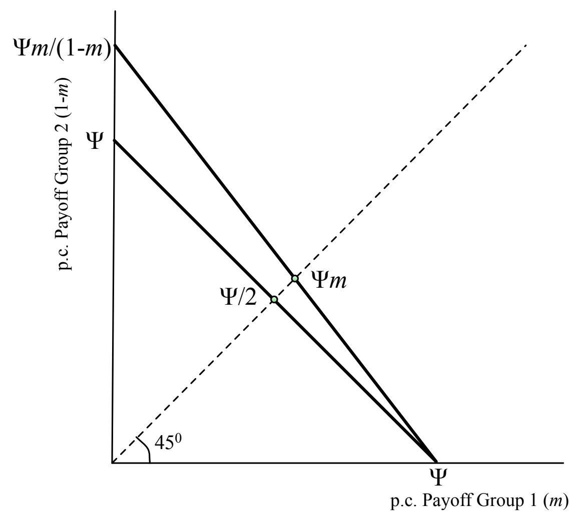


Figure 3. Transferable vs. Nontransferable Payoffs

Figure 3 illustrates the difference between classical and budget transferability when there are two disjoint groups. Under budget transferability, per-capita surplus is simply transferred one for one with every budget dollar transferred, *irrespective of group sizes*. An uniform peacetime allocation awards equal amounts of the budget to each group independent of size, resulting in a per-capita payoff of $\Psi/2$ for everyone. In contrast, in the classically transferable case, the *entire* budget goes to the larger group; say the one of size m in the Figure. That corner solution yields a payoff of 1 to each member of the larger group, and a payoff of 0 to each outsider. To achieve an uniform allocation, compensate the outsiders by having each "insider" make a financial transfer of $\Psi(1-m)$ to them, leaving each insider with a per-capita payoff of Ψm. The outsiders get a total of $\Psi(1-m)m$, which when divided among them, yields a per-capita payoff of Ψm as well. Note that in general, $\Psi m > \Psi/2$, unless the two groups are of exactly equal size.

When only budget transferability is available, society becomes more conflict-prone. This is not surprising, of course, as budget transferability limits the flexibility of peacetime allocations. We nevertheless illustrate this here in a simple exercise, for two reasons. First, we demonstrate that conflict is now a possibility *even* in the two-group case (recall that it is never a possibility with classical transferability; condition (8) cannot be satisfied). Second, the intuition that large groups are conflictual continues to be upheld with budget transferability.

Consider just two disjoint groups. To accommodate budget transferability, presume that a fraction $\sigma \in (0,1)$ of the budget can indeed be freely allocated using financial transfers, while the remainder can only be "transferred" by reallocating the budget devoted to "producing" those goods. In that case, the uniform peacetime payoff per person is given by

$$\Psi \left[\sigma m_1 + (1-\sigma)\frac{1}{2} \right], \tag{9}$$

where m_1, as before, is the size of the larger group. If only budget transferability is possible, this payoff drops to $\Psi/2$, while in the classical case with full transferability, it is Ψm_1.

Proposition 3. *Assume that the prize is public, and that there are two groups that partition society. Then there is* $m_1^* \in (0,1)$ *such that society is conflict-prone if and only if the larger group size* m_1 *exceeds* m_1^*.

Proof. Consider any conflict involving a Rebel of size m and the State of size $\overline{m} = 1 - m$. Then, using the expression (19) while noting that γ equals precisely 1 in the case at hand, the expected per-capita payoff to the Rebel is just $\Psi[km_1 + (1-k)m_1^2]$. It follows that conflict will occur if and only if

$$km_1 + (1-k)m_1^2 > \sigma m_1 + (1-\sigma)\frac{1}{2},$$

where m_1 is the larger group size in society. Define m_1^* by equality in the relationship above to complete the argument. ■

It is easy enough to conduct a similar analysis when there are n groups, with $n > 2$. Assume these groups are disjoint, and that *only* the budget can be transferred, so that $\sigma = 0$. Then it is easy to check that for any Rebel, conflict payoffs are given by (19), with the common value of γ equal to $(n-1)^{1/\alpha}$.[14] A necessary and sufficient condition for conflict is therefore

$$k\frac{m_1(n-1)^{1/\alpha}}{m_1(n-1)^{1/\alpha} + (1-m_1)} + (1-k)\left[\frac{m_1(n-1)^{1/\alpha}}{m_1(n-1)^{1/\alpha} + (1-m_1)}\right]^2 > \frac{1}{n},$$

and some straightforward but tedious calculations eventually reveal that

$$m_1 > \left[1 + (n-1)^{1/\alpha}\left\{\frac{(1+\alpha) - \sqrt{(\alpha-1)^2 + \frac{4\alpha}{n}}}{\sqrt{(\alpha-1)^2 + \frac{4\alpha}{n}} - (\alpha - 1)}\right\}\right]^{-1}. \tag{10}$$

For instance, when there are just two groups and the cost function is quadratic, then the Rebel needs to exceeds 61.8% of the population. When there are three groups and $\alpha = 1.2$, then the Rebel needs to exceed 39.7% of the population. We can use (10) to perform these calculations for any number of groups and any curvature of the cost function, but the point should be clear: it is large groups (typically but not always larger than the average) that pose a threat when the potential conflict is over public goods.

4. Actively Conflictual Societies

So far we have only been concerned with threats to uniform allocations. The theory developed thus far can be taken further, though from this point on we will not test the resulting implications. Say that a society is *actively conflictual* if every allocation, uniform or not, is blocked.

Whether or not a society can be actively conflictual is an interesting question because of the so-called Coase Theorem. Because conflict is costly, for each conflictual outcome there is a "peaceful" outcome that Pareto-dominates it, provided that appropriate Coaseian transfers are available. But is there *one* outcome that can *simultaneously* withstand all threats? The fact

[14] The reason is that the State also faces non transferability in the event of victory, so the per-capita value of victory for the State is $\Psi/(n-1)$.

that conflict is costly indicates that society is "superadditive," but the theorems of Bondareva (1963), Shapley (1967) and Scarf (1967) teach us that something more than mere superadditivity is needed to handle a multiplicity of potential demands from various blocking coalitions.

4.1. **Active Conflict with Private Payoffs.** When the prize is private, Proposition 1 shows that the presence of small minorities makes the society prone to conflict. Yet an uniform allocation is only one of several allocations society could choose in peacetime. After all, any of the conflict outcomes described in Proposition 1 can be Pareto-dominated by a suitably chosen allocation, because conflict is costly and therefore inefficient. (That allocation may or may not be uniform.) A key question, therefore, is whether there is some allocation that *simultaneously* avoids conflict from *all* potential Rebels. If the variety of potential threats is large relative to the degree of inefficiency, the answer is in the negative.

In the extension of this section, we assume that both for private as well as public prizes, there is a *collection* of potential Rebels who can each try to obtain control over the prize. To formalize the idea of a "variety of threats," say that a finite collection $\mathcal{C}$ of groups (or potential Rebels) is *balanced* if there is a set of weights in $[0, 1]$, $\{\lambda(G)\}_{G\in\mathcal{C}}$, such that

$$\sum_{G\in\mathcal{C}_i} \lambda(G) = 1 \text{ for every } i \text{ in society,} \tag{11}$$

where $\mathcal{C}_i$ is the subcollection of all groups for which i is a member.

What does balancedness mean? Essentially, it implies that it is hard to "buy off" small groups of individuals who are central to all potential conflicts. Balancedness assures us that there is no such "central group." For instance, suppose that $\mathcal{C}$ is fully described by any collection of potential Rebels that contain the special set of individuals $[0, 1/2]$. Then that collection is not balanced. For suppose we could find "balancing weights" $\{\lambda(G)\}$; then, in particular, (11) must hold for any $i \in [0, 1/2]$, but since i is contained in every $G \in \mathcal{C}$, this implies that the *entire* set of weights add to 1:

$$\sum_{G\in\mathcal{C}} \lambda(G) = 1.$$

Now pick any G' with $\lambda(G') > 0$. Because G' is a strict subset of $[0, 1]$, there is some individual $j \notin G'$. Given (12), it must be the case that

$$\sum_{G\in\mathcal{C}_j} \lambda(G) < 1,$$

which contradicts balancedness. Note how this unbalanced collection contains some distinguished group (in this example, $[0, 1/2]$) which is "over-represented" in the collection. In contrast, a balanced collection contains no "over-represented" group. For instance, any partition of $[0, 1]$ is a balanced collection (simply use $\lambda(G) = 1$ for all G and verify that the balancing condition is satisfied). Or, if $[0, 1]$ is the union of K equally-sized intervals of the form $[i/K, (i + 2)/K]$, for $i = 0, \ldots, K-1$, then the collection $\{[0, 2/K), [1/K, 3/K), [2/K, 4/K), \ldots, [(K-2)/K, 1), [(K-1)/K, 1/K)\}$ has "overlaps" but is also balanced.

We can now state:

Proposition 4. *Assume that the prize is private. Suppose that the collection of all potential Rebels includes a balanced collection $\mathcal{C}$, with each member of cardinality $m < m^*$, where m^* is given by Proposition 1.*

Then the society is actively conflictual.

Proof. Suppose that the conditions in the proposition are met, but that there is indeed an unblocked allocation $\mathbf{x}$. For every group $G \in \mathcal{C}$, we have

$$\int_{j \in G} x(j) \geq v[kp(m) + (1-k)p(m)^2] > vm. \tag{12}$$

Pick a collection of balancing weights $\{\lambda(G)\}_{G \in \mathcal{C}}$. Multiplying each side of (12) by $\lambda(G)$, and adding over all groups in $\mathcal{C}$, we see that

$$\sum_{G \in \mathcal{C}_j} \lambda(G) \int_{j \in G} x(j) > \sum_{G \in \mathcal{C}_j} vm\lambda(G).$$

Because $\{\lambda(G)\}_{G \in \mathcal{C}}$ are balanced weights, this implies

$$\int_j x(j) > v,$$

a contradiction. ■

In the light of our discussion following the definition of balancedness, this corollary applies:

Corollary 1. *Suppose that society can be partitioned into potential Rebels of size $m < m^*$. Then there is no allocation for society that is immune to conflict.*

The proof of this is immediate once we recognize that a partition of a society into groups is indeed a balanced collection.

At the heart of the argument for active conflict is the fact that when the prize is private, the complement of a minority Rebel — what we're calling the State — is generally weaker (in a per-capita sense) than the Rebel itself. In the case of private prizes considered in this section, that relative weakness is a manifestation of the Pareto-Olson thesis: that small groups contest the prize more vigorously (relative to their size) compared to large groups. It should be noted that our argument is strengthened further if a large defending coalition of groups can be further torn apart by subsequent bouts of conflict. That weakens further the resolve of the defendant to fight, knowing that its payoff from winning is even lower than what we've assumed here. We do not pursue such an extension in this paper.

Finally, the balancedness condition on potential Rebels, while sufficient, is not necessary. Sharper results are available. For instance, for groups that are smaller than the threshold m^*, extra per-capita surplus is available in the event of conflict.[15] For instance, suppose that the cost function is quadratic (so that $\alpha = 2$). It is then easy to verify that $m^* = 1/4$. However, groups of size

[15] At the same time, we should be careful not to take these assertions too literally, as the model ignores the fact that *some* minimum threshold size is probably needed to even pose a serious threat.

10% make a strict gain from blocking an uniform allocation. It is possible to check that if there are six such pairwise disjoint groups, conflict is inevitable regardless of the baseline allocation: no such allocation can be stable.

4.2. **Active Conflict with Public Payoffs.** A similar situation can occur with public prizes. Once again, we need to formulate the idea that the variety of potential threats is large, while avoiding the existence of common intersections that are pivotal for all threats, and so can be bought off.

Proposition 5. *Suppose that there exists a balanced collection of groups such that for the smallest group s in the collection,*

$$\frac{\eta_s(\eta_s + k)}{(\eta_s + 1)^2} > m_1, \tag{13}$$

where $\eta_s \equiv \left(\frac{m_s}{1-m_s}\right)^k / m_1^{1/\alpha}$. Then society is actively conflictual.

Proof. See Appendix. ■

We make two remarks on Proposition 5. First, as in the case of Proposition 4, a defending group is generally weaker (in a per-capita sense) than its complement, the Rebel. In the case of private prizes, that happened because of the Pareto-Olson effect. In the case of public goods, it happens because the State is heterogeneous (recall that the conflict condition can never be satisfied when there is a bilateral conflict across two homogeneous groups). Moreover, just as in the case of Proposition 4, the State can be weakened further by the threat of subsequent conflict, something we do not model here. That would only expand the set of scenarios under which active conflict is an outcome.

Second, one might ask if the condition of the proposition is too strong in the sense that it can never be satisfied. The following special case tells us that that isn't so. Suppose that society is *partitioned* into $n \geq 2$ groups, each of equal size. Observe that there is a unique value of n, call it $\hat{n}$, such that

$$(n-1)^{1-k} - 2 > (n-1)^k - kn \tag{14}$$

if and only if $n \geq \hat{n}$.[16] The partition is obviously a balanced collection, and condition (17) is the same for every element in it, and can be written as

$$\frac{1}{n} > \frac{1 - (n-1)^{1/\alpha} k}{[(n-1)^{1/\alpha} - 1]^2}, \tag{15}$$

because the (common) value of γ is given by $\gamma = (n-1)^{1/\alpha}$ for every group. It is easy to verify that (15) is the same condition as (8), and will therefore hold whenever the number of groups exceeds $\hat{n}$.[17]

[16]The left-hand side of this inequality is increasing in n, while the right hand side is decreasing in n, for all $n \geq 2$. To check the second claim, treat n as a continuous variable; then the derivative of the right-hand side of (14) is $k(n-1)^{k-1} - k$, which is strictly declining in n for $n \geq 2$.

[17]Cross-multiply the terms in (15) and divide through by $(n-1)^{1/\alpha}$ to arrive at (14).

Thus both conflict-proneness (in the face of uniform allocations) and active conflict (no matter what the allocation in place) are both possibilities. Moreover, we have seen that such conflicts are typically initiated by groups that are large, assuming that the baseline allocation is uniform.

4.3. **A Remark on the Salience of Ethnic Conflict.** We do not have a comprehensive theory of how certain classifications (religious, geographic, or ethnic) might acquire salience in conflict. Esteban and Ray (2008) pursue one line of research on this subject, which is that ethnic groupings permit each group to exploit the synergy of money and labor when engaging in conflict. The current exercise points to a different avenue for ethnic salience. Post-colonial societies have inherited or developed institutions — progressive taxation, land reform, public provision of education or health care — that are sensitive to threats along class lines. Such class-sensitive institutions are no coincidence, as the colonizing countries from which these newcomers have separated have had centuries of experience in developing those very institutions. But there are few analogous institutions for the differing fiscal treatment of *ethnic* groups. It is not that this cannot be done, or never has been done. It is just that such fiscal discrimination is generally difficult under a legal or constitutional umbrella. Therefore, one might conjecture that conflict organized along ethnic lines is a more likely outcome than conflict organized along class lines. Society has developed more institutions to take care of the latter, rather than the former. This dynamic of sluggish institutional adaptation may be at the heart of many conflictual societies, and it will be worth studying in future research.

5. Empirics

This section explores the relationship between group size, the nature of the payoffs, and conflict. Our theory implies that the impact of group size on conflict depends on the nature of the prize: smaller groups are more likely to be involved in conflict if the prize is private, while the size of the Rebel group is larger if the prize is public. There are several considerations that arise when using the data to address the theory. These include, but are not limited to, a suitable definition of "groups," as well as a classification of conflicts into their "private" and "public" payoff components. We also need to be careful about transplanting the "initiation of conflict" to the data.

5.1. **Taking the Theory to the Data.** The first empirical question is how to choose the social cleavage (or cleavages) that define potential Rebel groups. We settle for ethnicity, and study ethnic conflicts. Given that such conflicts account for between 50–75% of internal conflicts since 1945 (Fearon and Laitin 2003, Doyle and Sambanis 2006), this appears to be a natural and relatively tractable choice. As already discussed, our theory of "multiple threats" suggests channels that could account for the salience of ethnicity (as opposed to class) in conflict.

The second question has to do with the definition of a private goods conflict. Here we follow closely the guidelines of the theory. We do exactly as the model suggests and consider resources that are located on the homeland of each ethnic group. In our baseline specification this is oil, but we also consider other minerals as well as the size of the homeland itself. The presumption is that the State seeks to divide those resources more widely across the country, and the ethnic group in question can either accept the State-imposed status quo, or reject it.

The third issue is what constitutes a public goods conflict. This is a harder question and throughout the analysis we maintain both a narrow and a broad perspective. For the narrower perspective we focus on the seizure of political power at the Center. We proxy the payoff to that seizure using an *autocracy index* from Polity IV; see details below. The idea is that if the State is classified as autocratic to begin with, there will more to gain for an ethnic group by seizing the reins of that power. Or it may be that the disaffected who seize power simply want to get rid of the shackles of authoritarian rule and perhaps install a democratic government. Either interpretation will do.

The broader perspective is that all conflicts *not* having to do with resources on the ethnic homeland are conflicts over public goods in some shape or form. This may be overly inclusive a vision of public-goods conflict, while the view based on the autocracy index may seem too narrow, but it will at least be reassuring that both these perspectives yield similar results.

On the questions of public vs private goods conflicts, we should also note the obvious: that while the data are replete with conflicts over private and public payoffs, often there are mixtures of the two. For instance, even a conflict as seemingly primordial as Rwanda was permeated with economic looting, such as land grabs under the cover of ethnic violence. The Second Civil War in the Sudan is about different cultural and religious identities, but it is also — to some degree — about oil; so is the Chechnyan War. The Zimbabwean conflict is about identity and political power, but it is also about land, and so on.

The fourth question has to do with whether we study conflict *incidence* or *onset*. Briefly, a case of incidence records all conflict in a given time period, whether it is new or ongoing, while a case of onset records just the former. In our view, either approach can be defended, though in the case of incidence one should be careful to control for lagged conflict. Below, we take our baseline model to be one of incidence, though we explore variations that use onset (with similar results).

It should be noted that in all cases, the data we employ — a subset of the UCDP/PRIO Armed Conflict Dataset —records *only* conflicts between ethnic groups *and the State*. Below, we describe our empirical strategy in more detail. Our main results are presented in Sections 6 and 7. See Appendix B for detailed definitions of all the variables considered in the empirical analysis as well as a table of summary statistics.

5.2. **Data.** We have constructed a panel dataset at the ethnic group level with global coverage.[18] The dataset contains information for 145 countries and 1475 ethnic groups spanning the years 1960 to 2006.[19]

5.2.1. *Ethnic Groups.* We use the sample of ethnic groups from the dataset "Geo-Referencing of Ethnic Groups" (GREG); see Weidman, Rod and Cederman (2010). The GREG dataset provides detailed geographical location of ethnic groups for the whole world. This last feature enables us to merge with it other geo-referenced datasets needed for the computation of some of our key group-level variables. The GREG is based on the *Atlas Narodov Mira* or ANV (Bruk and

[18]This dataset is similar to that employed by Morelli and Rohner (2015) who consider similar sources for ethnic group location and oil fields.

[19]We focus on the post-1960 period as our data on ethnic group location and population are drawn from the start of the 1960s.

Apenchenko, 1964), which was created by Soviet ethnographers in the early 1960 with the aim of charting ethnic groups world wide. It provides information on the homelands of 929 groups and it employs a consistent classification of ethnicity with a uniform group list that is valid across state borders.[20] Most homelands are coded as pertaining to one group only, but in some instances up to three ethnic groups share the same territory.

The GREG extension of ANV permits us to create units that are group-country pairs: that is, we assign ethnic groups to countries depending on the land area occupied by them in each country.[21] When all is said and done, GREG contains a larger number of groups than alternative sources (such as the Geo-Ethnic Power Relations dataset) as it contains many small-language groups. There are 1475 distinct group-country pairs in the dataset, to be referred to from now on simply as "group." Our central variable, SIZE, is the size of the (country-specific) group relative to that of the population.[22]

The fact that GREG's settlement patterns — and our consequent classification of groups — are a snapshot from the late 1950s and early 1960s has advantages and disadvantages. On the negative side, settlement patterns may be outdated for some parts of the world. Also, as ethnic maps were chartered by Soviet ethnographers during the Cold War, the level of accuracy and resolution varies considerably for different regions in the world. On the positive side, it alleviates concerns of ethnic group location being endogenous to the conflicts we aim to explain.

5.2.2. *Conflict.* Data on group-level conflict has been taken from Cederman, Buhaug and Rod (2009), CBR henceforth.[23] We use two measures of conflict. Group-level conflict *incidence* is equal to 1 in a given year if that group is involved in an armed conflict against the state, resulting in more than 25 battle-related deaths in that year. Group-level conflict *onset* is equal to 1 in a given year if an armed conflict against the state resulting in more than 25 battle-related deaths *begins* in that year. For ongoing conflicts, onset is coded as missing. Our baseline specification uses conflict incidence.[24] We also show that our conclusions are robust to using onset.

5.2.3. *Prizes.* A key prediction of our theory is that the size of the group in conflict depends on whether the payoff is private or public. In order to test this hypothesis, proxies for the nature of the prize (or prizes) at stake are needed. To construct such proxies, we closely follow the approach in Esteban et al. (2012).

[20]The ANV actually contains information for 1248 groups, but 319 of them do not have any territorial basis.

[21]The definition of ethnic group is not clearly stated anywhere in the ANV so it is only possible to infer the coding criteria by comparison with existing data sources on ethnic groups. Fearon (2003) argues that the main criterion in the ANV for distinguishing between two groups is the historic origin of language.

[22]Population figures correspond to the early 60's, see Cederman, Buhaud and Rod (2009) for details.

[23]CBR use the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al. 2002) and check this list against previous sources that identify ethnic civil wars (such as Fearon and Laitin 2003, Licklider 1995 and Sambanis 2001). Ethnic conflicts are coded based on whether mobilization was shaped by ethnic affiliation. Once a list of plausible conflicts was established, CBR code the various groups involved in each case.

[24]In practice, conflict onset as defined by the PRIO threshold is far from a sharp concept. Before the threshold is crossed, we might have several manifestations of serious conflict (a breakdown in negotiations, an insurgency, a crackdown). Thus, a year of onset is arguably no different from a year of incidence, though to be sure, the factors that contribute to the outbreak of a conflict do not coincide with the ones that continue to feed it (Schneider and Wiesehomeier 2006). This is why we control for lagged conflict in our incidence regressions.

Private Prize. To obtain a proxy for the private payoff, we ask if the ethnic homeland is rich in natural resources. This is exactly in line with the theory. In our baseline specification we use oil abundance in the homeland as a proxy of "privateness". In the robustness check section we also consider mineral availability and land abundance (see Section 7.2).

Our baseline measure of group-level oil abundance, OIL, is computed as follows. First, geo-referenced information on the location of oil fields and associated discovery dates is obtained from PETRODATA (Lujala, Rod and Thieme, 2007). Next, we combine the information on group and oil location from GREG and PETRODATA, respectively, to construct maps of oil fields at the ethnic group level. Finally, OIL is computed as the log of the ethnic homeland area covered by oil (in thousands of square kilometres) times the international price of oil. Our results are robust to alternative ways of measuring oil abundance (see Section 7.2).

Notice that private prizes are firmly tied to ethnic homelands. As in the theory, implicit in this formulation is the idea that one ethnic group cannot directly reach out to seize resources located in another group's homeland. The State as a whole can, of course, attempt to redistribute the revenues from those resources over the country as a whole, or settle relatively abundant lands with other ethnicities. If violent conflict occurs in that process, our data will pick it up.

Public Prize. Our specific measure of publicness rests on the idea that there are large gains to seizing unconstrained power. Specifically, the more "autocratic" a country is, the less power is shared, the larger the number of citizens/groups that are excluded from power and consequently, the higher the valuation of controlling the State for larger groups in society. This may be because such groups are interested in seizing autocratic power themselves, or it may be because those groups want to install a democracy. In addition, as mentioned in Section 5.1, we also adopt a broader view of "public prizes" as pertaining to any situation in which the private prizes described above have been "netted out" or controlled for; more on this interpretation below.

Returning to the specific measure, we use the *autocracy index*, which is a composite measure from Polity IV.[25] The Polity IV manual summarizes the index thus: "[We] define [autocracy] operationally in terms of the presence of a distinctive set of political characteristics. In mature form, autocracies sharply restrict or suppress competitive political participation. Their chief executives are chosen in a regularized process of selection within the political elite, and once in office they exercise power with few institutional constraints . . . Our operational indicator of autocracy is derived from codings of the competitiveness of political participation, the regulation of participation, the openness and competitiveness of executive recruitment, and constraints on the chief executive."

We deliberately take this measure off the shelf so as to avoid any implication that the components or weights are chosen to suit our purpose. We are also aware that there are concerns of endogeneity: for instance, conflict can conceivably lead to changes in the autocracy index. Therefore, we *only* consider pre-sample values of the autocracy index (and in addition we control for past conflict in all our regressions). Specifically, our main "publicness" measure, AUTOC, is computed by averaging the values of the autocracy index from the end of the Second World War to

[25]This index is measured on a scale from 0 to 10, where 10 indicates the highest degree of autocracy, see Polity IV for details about its construction. We normalize this index to be between 0 and 1.

1960, which is the beginning of our sample.[26] The resulting measure is "assigned" to all the ethnic groups in the country, so that AUTOC is a time-invariant country-level index. This restriction is a necessary price to be paid for some acceptable degree of exogeneity: the variation of our resulting publicness measure is considerably smaller than that of the privateness measure, which is group specific and time-varying.

We check the robustness of our measure by considering three alternative proxies for the publicness index; see Section 7.3: an alternative index from Esteban, Mayoral, and Ray (2012), proxies for publicness based on religious freedom, and a group-level measure of exclusion from power. Our results are robust to using these alternative definitions.

5.2.4. *Additional Controls.* We also consider a number of control variables, both at the group and at the country level. Group-level controls have been obtained from Cederman et al. (2009) or have been directly computed from the GREG dataset. MOUNT is an index that captures the group's share of mountainous terrain. GROUPAREA is homeland area (in thousands of square kilometres). DISTCAP measures the group's distance to the country capital. GIP is one if the ethnic group is in power in a given country, lagged one year. SOILCONST is a measure of the limitations that the group's soil presents to agriculture. PARTITIONED is 1 if the group's homeland is located in two or more countries. PEACEYRS is the number of years since the last group-level onset and LAG is lagged conflict incidence. At the country level we control for the log of GDP per capita, lagged one year (GDP) and the log of total population (POP), also lagged one year. Both variables are taken from the Penn World Tables.

5.3. **Estimation.** We examine the effect of group size on conflict running variants of the following specification,

$$\text{(16)}\quad \text{INCIDENCE}_{c,g,t} = \beta_1 \text{SIZE}_{c,g} + \beta_2 \text{SIZE}_{c,g} \times \text{OIL}_{c,g,t} + \beta_3 \text{OIL}_{c,g,t} + \beta_4 \text{SIZE}_{c,g} \times \text{AUTOC}_c + X'_{c,g,t}\alpha + Y'_{c,t}\delta + Z'_c\gamma + W'_t\eta + \epsilon_{c,g,t},$$

for countries $c = 1, \ldots, C$, groups $g = 1, \ldots, G_c$, and dates $t = 1, \ldots, T$. Recall that OIL and AUTOC are our baseline measures of privateness and publicness, respectively, and their interactions with size are of particular interest. In addition, there are group- and country-level controls ($X_{c,g,t}$ and $Y_{c,t}$ respectively), a vector Z_c of country fixed effects and year dummies W_t.

Our theory predicts that β_2, the coefficient associated to SIZE $\times$ OIL, is negative, implying that smaller groups are more likely to be involved in conflict as oil in the homeland becomes more abundant. As for the public prize, recall from Section 5.1 that we adopt both a narrow and a broad view of "public prizes." The narrow view uses autocracy as a specific measure, and predicts that β_3, the coefficient associated to the interaction of group size and AUTOC, is expected to be positive, suggesting that the impact of group size on conflict increases as the public prize gets larger. The broad view is that *all* residual conflict after the effect of SIZE $\times$ OIL has been controlled for, is indeed conflict over a public prize. Under that view, it is to be expected that if

[26]For countries that didn't exist in 1960, the first non-missing value of the autocracy index is used instead.

$\beta_4 = 0$ is imposed, then $\beta_1 > 0$, assigning a positive role of group size to the precipitation of such conflicts.

Throughout, we have country fixed effects. Identification for the interaction term SIZE× OIL is achieved both because we have variation in ethnic groups within countries — so that size varies — and intertemporal variation in oil prices or in known reserves. However, the only source of variation for the interaction term SIZE × AUTOC is changes in ethnic groups within countries, because AUTOC is a *country-level*, time-invariant indicator. This is the main reason why we do not use group fixed effects, though in one version (see Section 7.4.1) we explore this case, as group-level variation in SIZE× OIL is still possible through the OIL component.

We estimate equation (16) by OLS. The reason for fitting a linear probability model (rather than a non-linear specification, such as probit or logit) is that our key variables are interactions and interpreting them in nonlinear models isn't straightfoward, as Ai and Norton (2003) point out.[27] Thus, for the sake of simplicity we will focus on linear specifications. In Section 7.5, however, we also consider estimation in nonlinear models. Robust standard errors, clustered at the group level — that is, for each ethnic-group/country pair — have been computed in all cases.[28] We also show that our results are robust to clustering errors at the country level and at the country and ethnic homeland level (two-way clustering), where the latter considers all the territories occupied by the same group (even if they belong to different countries); see C.3.

6. RESULTS

Table 1 reports our baseline results. The dependent variable is conflict incidence. Each column reports on a different linear probability specification, all with lagged conflict and country and time fixed effects. Column 1 regresses INCIDENCE on only two variables: group size (SIZE) and group-level oil abundance (OIL). The coefficient of SIZE is small and not significant. This is precisely what the theory would lead us to expect, as it predicts that the *unconditional* effect of group size on conflict is ambiguous. On the other hand, the abundance of oil in the ethnic homeland is positively associated with conflict incidence.

Column 2 introduces the interaction of SIZE and OIL. The coefficient of the interaction term is negative and significant, as predicted by the theory. Column 3 (4) adds group (country)-level controls and analogous results are obtained.

Column 5 (henceforth our baseline specification) is similar to column 4 but also includes in the regression SIZE and its interaction with AUTOC. Observe that AUTOC cannot be introduced as an

[27]In linear models, the coefficient of the interaction term has a direct interpretation, as it is just the value of the cross derivative of the dependent variable with respect to the variables in the interaction. However, this logic does not extend to nonlinear models: the cross derivative in this case is a more complicated object. As shown by Ai and Norton (2003), its value depends on all the covariates of the model and the sign does not necessarily coincide with the sign of the coefficient of the interaction, see Appendix C.1 for a more detailed discussion.

[28]An alternative would be to cluster the standard errors at the country rather than at the group level. However, the asymptotic validity of the clustering methods relies on the number of clusters going to infinity with the sample size. As the number of countries compared to the number of groups is small in our case, this type of clustering is likely to perform poorly (Wooldridge 2003). Clustering standard errors at the country rather than at the group level doesn't modify our conclusions, see Table C2 in Appendix C.

	Dependent Variable: **Conflict Incidence**				
	[1]	[2]	[3]	[4]	[5]
SIZE	-0.002	0.003	0.007***	0.007***	0.003
	(0.307)	(0.101)	(0.001)	(0.001)	(0.156)
OIL	0.448**	0.684***	0.830***	0.795***	0.762**
	(0.040)	(0.009)	(0.002)	(0.008)	(0.010)
SIZE×OIL		-1.363***	-1.528***	-1.521***	-1.390***
		(0.000)	(0.000)	(0.000)	(0.000)
SIZE× AUTOC					0.009**
					(0.015)
GIP			-0.003**	-0.003*	-0.003*
			(0.033)	(0.057)	(0.057)
GROUPAREA			0.000	0.000	0.000
			(0.369)	(0.214)	(0.219)
SOILCONST			-0.001*	-0.000	-0.000
			(0.097)	(0.518)	(0.472)
DISTCAP			0.001***	0.002***	0.002***
			(0.000)	(0.000)	(0.000)
MOUNT			0.002*	0.002	0.002
			(0.080)	(0.111)	(0.130)
PARTITIONED			-0.001	-0.001	-0.001
			(0.553)	(0.288)	(0.243)
GDP				0.001	0.003***
				(0.140)	(0.006)
POP				0.001	0.001
				(0.556)	(0.710)
LAG	0.895***	0.895***	0.894***	0.893***	0.898***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-0.002	-0.005***	-0.009***	-0.034	-0.041
	(0.207)	(0.006)	(0.000)	(0.411)	(0.319)
R^2	0.844	0.844	0.844	0.846	0.851
Obs	64839	64839	64839	57559	55383
R^2	0.844	0.844	0.844	0.846	0.851
Obs	64839	64839	64839	57559	55383

Table 1. Group Size and Conflict: Baseline. This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain year dummies and country fixed effects. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

independent regressor in the equation, as it is measured at the country level and is time-invariant. Thus, it is included in the country fixed effects. *Both* interactions return the predicted signs and they are significant.[29]

[29] As mentioned before, standard errors have been clustered at the country-group level (rather than at the country level) because the asymptotic validity of the clustering methods relies on the number of clusters going to infinity with the sample size and, since the number of countries compared to the number of groups is small in our case, this type of clustering is likely to perform poorly (Wooldridge 2003). For the sake of robustness, we have re-estimated Table 1 clustering standard errors at the country level. We've also re-estimated it using two-way clustering, with errors clustered at the group (as opposed to country-group) *and* country level. Our conclusions remain identical; see Table C2 in Appendix C.

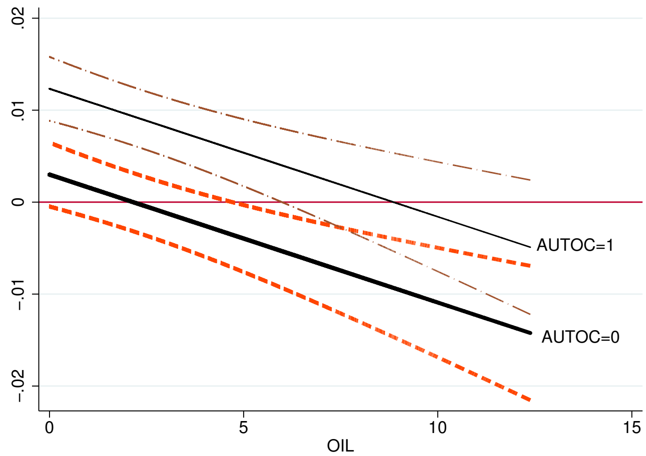


Figure 4. Marginal effect of group size on conflict *incidence* as a function of OIL
This graph depicts the marginal effect of group size on conflict incidence as a function of OIL for two different values of AUTOC: AUTOC=0 (bottom solid line) and AUTOC= 1 (top solid line). Confidence bands at the 95% level are also depicted. Estimates from Table 1 (column 5) have been employed to compute the estimates.

This baseline table — as well as an overwhelming majority of the variations to come — contains an additional result. Observe that once the interaction term SIZE× OIL is in the place, the coefficient of SIZE *alone* now captures the effect of group size on conflict "in the absence of a private prize." It is positive and significant, suggesting that when there is no oil larger groups are more likely to be involved in conflict. This effect speaks to the broader view of public payoffs that we've described in Section 5.1 and elsewhere. If we provisionally agree that all size effects, other than those explicitly captured in our size-resource interaction, pertain to public conflicts, then we have additional evidence that group size has a positive effect on conflicts with public prizes.

It is also of interest that SIZE ceases to be significant when the interaction between size and the public prize is introduced in the regression. That suggests that AUTOC and OIL are suitable proxies for the public and private prizes, respectively. While this auxiliary result does vary somewhat across specifications (see, for instance, the last column of Table 2), it generally appears to hold in the bulk of the variations.

These results clearly suggest — in line with the theory — that the effect of group size on conflict critically depends on the nature of the potential payoffs. Figure 4 plots the marginal effect of SIZE on INCIDENCE computed using the estimates from Column 5 in Table 1. The marginal effect is a function of both OIL and AUTOC, and the plot displays the marginal effect as a function of OIL (in the X axis), for the minimum and maximum values of AUTOC (AUTOC=$\{0, 1\}$).[30] The thick (thin) solid line corresponds to the marginal effect of group size on conflict as a function of OIL for AUTOC=0 (AUTOC=1). The dashed lines represent 95% confidence bands.

[30] The marginal effect is simply obtained by differentiating equation (16) with respect to SIZE and inserting the estimates from Column 8 into the resulting expression.

Dependent Variable: **Conflict Onset**					
	[1]	[2]	[3]	[4]	[5]
SIZE	-0.001	0.003**	0.005***	0.005***	0.003*
	(0.333)	(0.025)	(0.001)	(0.001)	(0.053)
OIL	0.652***	0.870***	0.966***	0.937***	0.957***
	(0.002)	(0.001)	(0.000)	(0.001)	(0.001)
SIZE×OIL		-1.221***	-1.171***	-1.149***	-1.079***
		(0.000)	(0.000)	(0.000)	(0.000)
SIZE× AUTOC					0.005*
					(0.069)
GIP			-0.002*	-0.002*	-0.002*
			(0.076)	(0.078)	(0.092)
GROUPAREA			-0.000	-0.000	-0.000
			(0.376)	(0.659)	(0.613)
SOILCONST			-0.000	-0.000	-0.000
			(0.102)	(0.479)	(0.466)
DISTCAP			0.001***	0.001***	0.001***
			(0.001)	(0.003)	(0.004)
MOUNT			0.002**	0.002**	0.002*
			(0.017)	(0.048)	(0.055)
PARTITIONED			-0.000	-0.001	-0.001
			(0.716)	(0.407)	(0.328)
GDP				0.001	0.002**
				(0.301)	(0.045)
POP				0.002	0.002
				(0.263)	(0.237)
PEACEYRS	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	0.070***	0.067***	0.012***	0.009	-0.012
	(0.000)	(0.000)	(0.001)	(0.795)	(0.618)
R^2	0.030	0.031	0.031	0.033	0.034
Obs	63187	63187	62762	55611	53466

Table 2. Group Size and Conflict: Onset. This table regresses conflict onset on group size and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain year dummies and country fixed effects. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Figure 4 shows that the marginal effect of size can be negative or positive, depending on the values of the public and private payoffs. For a small value of AUTOC and moderate or large values of OIL, the effect of an increase in group size has a negative and significant effect on conflict incidence. The opposite is true when AUTOC is high and OIL is small: in this case the marginal effect of SIZE is positive and significant. However, it is not significantly different from zero when either both prizes are small or when both are large.

Finally, we provide a sense of the magnitudes of the coefficients. As these depend on the values of AUTOC and OIL, we provide a couple of examples. For AUTOC $= 0$ and a high value of oil (at the 95th percentile) an increase of one standard deviation in SIZE decreases the unconditional probability of conflict incidence by 4.2%. Similarly, if OIL $= 0$ and AUTOC is high ($= 1$), an increase of one standard deviation in SIZE increases the probability of conflict by 9.5%.

7. VARIATIONS

We now examine the robustness of the baseline results in Table 1. We consider (i) alternative measures of conflict, (ii) alternative ways of measuring the private prize, (iii) different proxies for the public prize, (iv) models with group fixed effects, (v) methods to assess the importance of the omitted variable bias, (vi) alternative estimation strategies (logit), and (vii) the possibility that groups form coalitions. We include further variations in the Online Appendix (Appendix C). Specifically, (viii) we cluster errors at the country and at the country and group level (two-way clustering), (ix) we consider robustness to dropping different regions of the world and (x) we consider the potential confounding role of ethnic fractionalization and polarization.

7.1. **Alternative Measures of Conflict.** Table 2 replicates Table 1 using conflict *onset* as dependent variable. Qualitatively, the results are very similar to those described above. The interactions of group size and the publicness/privateness indicators have the predicted sign and are highly significant. Quantitatively, the (relative) magnitude of the effect is larger in the onset than in the incidence regressions. Using the same examples as before, for AUTOC $= 0$ and a high value of oil (at the 95th percentile), an increase of one standard deviation in SIZE decreases the unconditional probability of conflict onset by 23.2%. Similarly, if OIL $= 0$ and AUTOC is high (=1), an increase of one standard deviation in SIZE increases the unconditional probability of conflict onset by 69.8%. Note again the telling insignificance of size when entered without any interactions at all, but also the reversal of that null result — with a positive and significant coefficient on size alone — once the size-resource interaction term is in place. Again, as already discussed, this is supportive of the broader view on public conflicts and the positive effect of group size on such conflicts.

7.2. **Alternative Ways of Constructing the Private Prize.** We've also considered alternative ways of constructing the privateness index. Columns 1 to 4 in Table 3 use alternative definitions of oil abundance. Columns 1 and 2 use OIL(AREA) which is defined as the log of the homeland area covered by oil. The difference between columns 1 and 2 is that the former doesn't include the interaction between SIZE and the public payoff, AUTOC. The results are very similar to those obtained in the baseline specifications. Columns 3 and 4 use relative, as opposed to absolute, values of oil. The variable OIL(SHARE) is defined as the *share* of the ethnic homeland covered by oil. The results are also similar as above. Columns 5 and 6 use a different proxy of privateness, as they focus on land availability. The variable AREA(SHARE) measures the share of the ethnic homeland area as a fraction of the total area of the country. The idea is that in case of conflict the available land can be seen as a private payoff, whose valuation will increase if land is relatively scarce in the rest of the country. Columns 5 and 6 show that ethnic groups whose homelands occupy a large share of the total area of the country are more likely to be involved in conflict. In addition, as predicted by the theory, the interaction of SIZE and AREA(SHARE) is negative and significant, suggesting that small groups are more likely to be involved in conflict as the value of AREA(SHARE) increases.

	Dependent Variable: **Conflict Incidence**					
	[1]	[2]	[3]	[4]	[5]	[6]
SIZE	0.006***	0.002	0.005***	0.001	0.018***	0.015***
	(0.004)	(0.366)	(0.009)	(0.683)	(0.003)	(0.006)
OIL(AREA)	0.002**	0.002**				
	(0.011)	(0.017)				
OIL(SHARE)			0.010*	0.010*		
			(0.082)	(0.093)		
AREA(SHARE)					0.022**	0.021**
					(0.030)	(0.041)
SIZE× OIL(AREA)	-0.003***	-0.003***				
	(0.001)	(0.003)				
SIZE× OIL(SHARE)			-0.021**	-0.017*		
			(0.018)	(0.055)		
SIZE×AREA(SHARE)					-0.042***	-0.041***
					(0.000)	(0.000)
SIZE× AUTOC		0.009**		0.010***		0.007*
		(0.016)		(0.009)		(0.058)
GIP	-0.003*	-0.004*	-0.003*	-0.003*	-0.005***	-0.005***
	(0.054)	(0.054)	(0.077)	(0.075)	(0.008)	(0.009)
GDP	0.001	0.003***	0.001	0.003***	0.001	0.003***
	(0.122)	(0.005)	(0.123)	(0.005)	(0.135)	(0.005)
POP	0.001	0.001	0.001	0.001	0.001	0.001
	(0.580)	(0.743)	(0.589)	(0.754)	(0.597)	(0.764)
GROUPAREA	0.000	0.000	0.000	0.000		
	(0.497)	(0.508)	(0.222)	(0.155)		
SOILCONST	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.501)	(0.471)	(0.897)	(0.834)	(0.532)	(0.493)
DISTCAP	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MOUNT	0.002	0.002	0.002	0.002	0.002	0.002
	(0.121)	(0.143)	(0.152)	(0.169)	(0.161)	(0.202)
PARTITIONED	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.309)	(0.262)	(0.377)	(0.325)	(0.293)	(0.246)
LAG	0.894***	0.898***	0.894***	0.898***	0.893***	0.898***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-0.034	-0.041	-0.036	-0.043	-0.003	-0.002
	(0.409)	(0.321)	(0.381)	(0.299)	(0.945)	(0.959)
R^2	0.846	0.851	0.846	0.851	0.846	0.851
Obs	57559	55383	57559	55383	56756	54580

Table 3. Variations: Alternative Private Prize Specifications. This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables. Columns 1–4 use alternative oil-based measures of privateness, and Columns 5–6 use land-based measures, as described in the text. All columns contain country fixed effects and year dummies, and have been estimated by OLS. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 4 considers mineral availability in the ethnic homeland as a proxy for "privateness." We use geo-referenced data on the location of mining activities around the world since 1980.[31] Since the mineral data starts in 1980, we use $\text{AUTOC}_{1960-80}$ (defined as the average between 1960 and 1980 of the Polity IV autocracy index) as a proxy for the public prize. For each year and mine, we know whether that mine is active or not, and the specific minerals produced by it. As in Berman et al (2015), we focus on 13 minerals for which we have world price data,[32] which

[31]The source is the *Raw Material Data* (IntierraRMG, 2015).

[32]These are Bauxite, Coal, Copper, Diamond, Gold, Iron, Lead, Nickel, Platinum, Phospate, Silver, Tin and Zinc.

	Dependent Variable: **Conflict Incidence**							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
SIZE	0.007**	0.003	0.008**	0.004	0.007**	0.003	0.008**	0.004
	(0.020)	(0.362)	(0.015)	(0.281)	(0.022)	(0.392)	(0.016)	(0.303)
MINES	0.000	0.000						
	(0.800)	(0.849)						
SIZE× MINES	-0.002**	-0.001**						
	(0.021)	(0.048)						
MINES(+OIL)			0.000	0.000				
			(0.561)	(0.602)				
SIZE× MINES+OIL			-0.002**	-0.002**				
			(0.012)	(0.029)				
MINES(UNWEIGH.)					0.000	0.000		
					(0.831)	(0.877)		
SIZE× MINES(UNWEIGH.)					-0.001**	-0.001*		
					(0.022)	(0.055)		
MINES+OIL(UNWEIGH.)							0.000	0.000
							(0.594)	(0.632)
SIZE× MINES+OIL(UNWEIGH.)							-0.002**	-0.001**
							(0.012)	(0.033)
SIZE× AUTOC$_{1960-80}$		0.009**		0.009**		0.009**		0.008**
		(0.027)		(0.034)		(0.028)		(0.036)
GIP	-0.002	-0.002	-0.002	-0.002	-0.001	-0.002	-0.002	-0.002
	(0.512)	(0.460)	(0.503)	(0.454)	(0.516)	(0.464)	(0.508)	(0.458)
GDP	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	(0.103)	(0.104)	(0.103)	(0.104)	(0.103)	(0.104)	(0.103)	(0.104)
POP	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.601)	(0.595)	(0.603)	(0.596)	(0.595)	(0.589)	(0.596)	(0.590)
GROUPAREA	0.002*	0.002*	0.002*	0.002	0.002*	0.002	0.002*	0.002
	(0.070)	(0.098)	(0.072)	(0.103)	(0.070)	(0.101)	(0.070)	(0.104)
SOILCONST	-0.000	-0.001	-0.001	-0.001	-0.000	-0.001	-0.001	-0.001
	(0.261)	(0.246)	(0.219)	(0.207)	(0.266)	(0.252)	(0.224)	(0.212)
DISTCAP	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
MOUNT	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	(0.358)	(0.338)	(0.358)	(0.338)	(0.358)	(0.338)	(0.358)	(0.339)
PARTITIONED	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.671)	(0.646)	(0.655)	(0.631)	(0.673)	(0.648)	(0.658)	(0.633)
LAG	0.886***	0.886***	0.886***	0.886***	0.886***	0.886***	0.886***	0.886***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	0.007	0.007	0.006	0.006	0.008	0.008	0.007	0.007
	(0.933)	(0.934)	(0.942)	(0.941)	(0.926)	(0.927)	(0.934)	(0.934)
R^2	0.836	0.836	0.836	0.836	0.836	0.836	0.836	0.836
Obs	35265	34887	35265	34887	35265	34887	35265	34887

Table 4. Variations: Alternative Private Prize Specifications With Minerals. This table regresses conflict incidence on group size and indices of private and public prizes. Different specifications using minerals are considered for the private prize. All regressions contain year dummies and country fixed effects, and have been estimated by OLS. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. AUTOC$_{1960-80}$ has been computed as the average from 1960 to 1980 of the autocracy index (Polity IV). The coefficient of GROUPAREA has been multiplied by 1000. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

we take from the World Bank's commodity price database. We create two types of indices, one that includes information on mineral prices and another that doesn't. They are constructed as follows: for each group, year and mineral, we create a dummy variable that is one if the group has at least one active mine of that mineral. The variable MINES(UNWEIGHTED) is computed by simply adding up the resulting dummies for each group and year. To introduce information on mineral prices, we multiply each of the mineral dummies by (the log of) its international price, normalized by (the log of the) price in 1980 (the year when the mineral data starts). The variable

MINES is constructed as the sum of the resulting quantities for each group and year. Finally, MINES+OIL(UNWEIGHTED) and MINES+OIL are constructed in an analogous way except that they also consider oil availability. Table 4 presents the results obtained using the above-defined measures as a proxy for privateness. In general our conclusions remain unchanged. That includes the positive effect of group size once the private interaction term is accounted for, which continues to support our broader perspective on public payoffs and the hypothesis that group size is positively related to such conflicts.

7.3. **Alternative Ways of Constructing the Public Prize.** Now we turn to variations on the public prize. Our baseline measure of publicness, AUTOC, is based on the pre-1960 values of the Polity IV autocracy index. However, many countries in our sample (around a quarter of them) didn't exist before 1960. In those cases, AUTOC is computed as the first non-missing observation of the autocracy index. In order to have a more homogeneous definition for all countries, we have considered an alternative measure computed as the average from 1960 to 1980 of the Polity IV autocracy index (that we denote as $\text{AUTOC}_{1960-80}$). Then, this index has been employed in regressions that only consider the second half of the sample (from 1980 onwards), so that reverse causality is less of a concern. Column 1 in Table 5 is similar to our baseline specification (column 8 in Table 1) but uses $\text{AUTOC}_{1960-80}$ as a proxy for publicness and only post-1980 observations enter the regression. This column shows that the interaction of SIZE and $\text{AUTOC}_{1960-80}$ is positive and significant while that of SIZE and OIL is still negative and significant. Moreover, the magnitude of the coefficients associated to these interactions and their significance level remain very stable when only the second half of the sample is employed.

Another potential concern is the different level of variation of our privateness and publicness proxies. Whereas AUTOC is time-invariant and defined at the country-level, the privateness variables are group-specific and time-varying. To account for this asymmetry, we have also considered a group-level measure of publicness, based on whether the group is excluded from State power. We employ three versions of this measure, one that varies over time and another two that are based on pre-sample averages and, therefore, are time-invariant. The variable EXCLUDED is 1 for group i if that group was excluded from power the previous year (i.e., EXCLUDED is simply 1-GIP). The variables $\text{EXCLUDED}_{1945-60}$ and $\text{EXCLUDED}_{1960-80}$ are computed by averaging the 1945–1960 and the 1960–80 values of EXCLUDED, respectively. Column 2 in Table 5 shows that the variable EXCLUDED is positive and significant, which indicates that excluded groups are more likely to be involved in conflict. Column 3 adds the interaction of SIZE and EXCLUDED to the specification in column 2. The coefficient of this interaction term is positive and significant (while now the variable EXCLUDED is insignificant), suggesting that excluded groups are more likely to be involved in conflict when they are large. Column 4 considers the variable $\text{EXCLUDED}_{1945-60}$ and shows that its interaction with SIZE has the expected positive sign but it is not significant (p-value is .15). The lack of significance of this term can be due to many reasons, one of them being that the 1945–1960 average of EXCLUDED can be a poor proxy for the current power status of the group. The variable $\text{EXCLUDED}_{1960-80}$, being more correlated with EXCLUDED, could be a better proxy for it. Column 5 uses $\text{EXCLUDED}_{1960-80}$ in regressions that only consider post 1980 observations. In this case, the coefficient of its interaction with SIZE is positive and highly significant, in line with our hypothesis that large excluded groups are more likely to be involved in conflict.

	Dependent Variable: **Conflict Incidence**						
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
SIZE	-0.000	0.007***	0.003	0.004	0.001	0.005**	-0.001
	(0.985)	(0.001)	(0.359)	(0.193)	(0.836)	(0.013)	(0.910)
OIL	0.695**	0.795***	0.760**	0.777***	0.719**	0.790***	1.162**
	(0.039)	(0.008)	(0.010)	(0.009)	(0.032)	(0.008)	(0.025)
SIZE× OIL	-1.217**	-1.521***	-1.371***	-1.555***	-1.143**	-1.369***	-2.138***
	(0.012)	(0.000)	(0.001)	(0.000)	(0.016)	(0.000)	(0.002)
SIZE× $\text{AUTOC}_{1960-80}$	0.008**						
	(0.039)						
EXCLUDED		0.003*	0.002				
		(0.057)	(0.354)				
SIZE× EXCLUDED			0.008*				
			(0.067)				
$\text{EXCLUDED}_{1945-60}$				0.002			
				(0.363)			
SIZE× $\text{EXCLUDED}_{1945-60}$				0.005			
				(0.148)			
$\text{EXCLUDED}_{1960-80}$					0.002		
					(0.471)		
SIZE× $\text{EXCLUDED}_{1960-80}$					0.012**		
					(0.015)		
SIZE× PUB(EMR)						0.009***	
						(0.002)	
GIP						-0.005***	-0.007**
						(0.003)	(0.050)
RELIGFREEDOM							0.043***
							(0.007)
SIZE×RELIGFREEDOM							0.021*
							(0.086)
GDP	0.003	0.001	0.001	0.001	0.003	0.001	0.004
	(0.106)	(0.140)	(0.134)	(0.137)	(0.106)	(0.145)	(0.399)
POP	-0.003	0.001	0.001	0.001	-0.003	0.001	-0.009
	(0.612)	(0.556)	(0.546)	(0.538)	(0.607)	(0.563)	(0.368)
GROUPAREA	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.652)	(0.214)	(0.209)	(0.263)	(0.523)	(0.137)	(0.894)
SOILCONST	-0.001	-0.000	-0.000	-0.000	-0.001	-0.000	-0.001
	(0.162)	(0.518)	(0.602)	(0.576)	(0.206)	(0.400)	(0.218)
DISTCAP	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.002)
MOUNT	0.002	0.002	0.002	0.002	0.002	0.002	0.000
	(0.202)	(0.111)	(0.111)	(0.107)	(0.310)	(0.112)	(0.926)
PARTITIONED	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.287)	(0.288)	(0.271)	(0.246)	(0.586)	(0.262)	(0.679)
LAG	0.886***	0.893***	0.893***	0.893***	0.885***	0.893***	0.832***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	0.008	-0.037	-0.035	-0.034	0.007	-0.033	0.055
	(0.926)	(0.362)	(0.394)	(0.395)	(0.948)	(0.418)	(0.778)
R^2	0.836	0.846	0.846	0.846	0.836	0.846	0.763
Obs	34887	57559	57559	57559	34965	57559	22166

Table 5. Variations: Alternative Specifications of the Public Prize. This table regresses conflict incidence on group size and indices of private and public prizes. Alternative specifications are considered for the public prize. All regressions contain year dummies and country fixed effects, and have been estimated by OLS. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Column 6 in Table 5 use the measure of publicness employed in EMR (2013); call it PUB(EMR). The index PUB(EMR) is a country average of several indicators of autocracy (such as the autocracy index from Polity IV, as well as Civil Rights and Political Freedom indicators from Freedom House), see EMR for details. Using PUB(EMR) instead of AUTOC yields very similar results.

Finally, Column 7 looks at an alternative measure of publicness that is based on religious freedom. The idea is that religious extremists will see existing religious freedoms as a space to be seized — as a *prize*, in short — as they would like society to behave according to their religious rules. To construct an index that reflects the lack of religious restrictions, we consider data from the *Religion and State Project* assembled by ARDA (`http://www.thearda.com/ras/`). This dataset starts in 1990 and provides detailed codings on several aspects of the government activities with regard to religion that are encompassed by the concepts of separation of religion and state and government involvement in religion. The variable RELIGFREEDOM measures the extent to which, in practice, a state is willing to restrict some or all religions. High values of this measure reflect a higher degree of religious freedom. [33] Column 6 shows that RELIGFREEDOM is positively associated with conflict. In addition, the interaction between SIZE and RELIGFREEDOM is positive and significant, suggesting that larger groups are more likely to fight for religious motives.

7.4. **Omitted Variable Bias.** Our baseline specification contains country and year fixed effects and also controls for several group-level characteristics. However, there is invariably the concern that some unobserved group-level characteristic might bias our results. To address this concern, we follow two strategies. First, we consider models with group fixed effects. Second, we follow Oster (2016) and assess the likelihood that our observed effect is due to selection bias.

7.4.1. *Group Fixed Effects.* The reason why we did not use group fixed effects in the first place is that we need variation in group sizes in order to identify the effect of SIZE× AUTOC, given that AUTOC is already subsumed in the country-fixed effects. With group fixed effects, all time-invariant controls drop out from the regression, including two of our key variables (SIZE and SIZE× AUTOC). Nevertheless, it is still possible to test one of the two key hypotheses, that pertaining to SIZE× OIL. Columns 1– 3 in Table 6 do just that. Note that the three columns contain group fixed effects but are still different, because Column 1 excludes lagged conflict, while Columns 2 and 3 include this variable and are estimated by OLS and system GMM (Blundell and Bond 1998). In all cases, the interaction of SIZE and OIL remains negative and significant.

7.4.2. *Assessing the Importance of the Omitted Variable Bias.* Despite our attempts to control for a large number of potential confounders, we still cannot completely rule out the possibility that unobserved variables are biasing our results. However, it is possible to assess the likelihood that our observed effect is solely due to selection bias. To that effect, we apply a technique recently developed by Oster (2016) which builds on the work by Altonji, Elder and Taber (2005) and Bellows and Miguel (2008). This method allows us to assess how strong the correlation between the unobservables and the key independent variables has to be (relative to the correlation between the observable covariates and the key independent variables) in order to explain away the observed result. If the set of observed controls is representative of all possible controls, then

[33] RELIGFREEDOM is coded on the following scale: 1. All (other) religions are illegal; 2. Some (other) religions or atheism are illegal; 3. No religions are illegal but some or all (other) religions have legal limitations placed upon them; 4. No religions are illegal but some or all (other) religions have practical limitations placed upon them; 5. No religions are illegal and no limitations are places on them but some religions have benefits not given to others due to some form of official recognition or status not given to all religions; 6. No (other) religions are illegal and there are no significant restrictions on minority religions.

	Dependent Variable: **Conflict Incidence**				
	[1]	[2]	[3]	[4]	[5]
OIL	0.006	0.001	-0.000	0.244***	0.165***
	(0.194)	(0.293)	(0.628)	(0.000)	(0.000)
SIZE× OIL	-0.011*	-0.003*	-0.001***	-0.838***	
	(0.060)	(0.100)	(0.008)	(0.000)	
SIZE× AUTOC					2.370*
					(0.073)
GIP	-0.032	-0.009	-0.000	-0.409	-0.454
	(0.147)	(0.181)	(0.928)	(0.106)	(0.106)
GDP	-0.006	0.001	0.004**	0.376**	0.628***
	(0.192)	(0.544)	(0.020)	(0.024)	(0.001)
POP	0.016	0.003	-0.000	1.772**	1.471*
	(0.157)	(0.340)	(0.618)	(0.039)	(0.074)
SIZE				2.339***	-0.161
				(0.000)	(0.843)
GROUPAREA				0.000	-0.000
				(0.523)	(0.397)
SOILCONST				-0.193	-0.067
				(0.167)	(0.610)
DISTCAP				0.001***	0.001***
				(0.000)	(0.001)
MOUNT				0.613***	0.472*
				(0.009)	(0.052)
PARTITIONED				-0.150	-0.185
				(0.323)	(0.258)
LAG		0.797***	0.919***	7.270***	7.383***
		(0.000)	(0.000)	(0.000)	(0.000)
c	-0.213	-0.051	-0.031**	-46.125***	-42.203***
	(0.317)	(0.317)	(0.039)	(0.007)	(0.009)
GFE/CFE	GFE	GFE	GFE	CFE	CFE
(Pseudo) R^2	0.012	0.640	–	0.8179	0.8210
Obs	57559	57559	57559	27344	25495

Table 6. Variations: Group Fixed Effects and Nonlinear Models. This table regresses conflict incidence on group size and indices of private and public prizes. All regressions contain year dummies. GFE (CFE) denotes group (country) fixed effects. Columns 1 and 2 have been estimated by OLS and Column 3 by system GMM (Blundell and Bond, 1998). Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

a large value of the ratio of these correlations suggests that it is implausible that omitted variable bias explains away the entire effect. Altonji et al. (2005) and Oster (2016) suggest the use of a cutoff value of 1 for the ratio of correlations. This value of the ratio means that the unobservables would need to be at least as important as the observables to produce a treatment effect of zero. One reason to favor this cutoff value is that researchers typically choose the controls they believe ex ante are the most important (Angrist and Pischke, 2010) and thus situations where the effect of the unobservables is larger than that of the controls are deemed unlikely.

In practice, the ratio of correlations is computed as a function of the coefficient of the variable of interest estimated in a full model with all controls, the same coefficient obtained in a restricted model with no (or few) controls and the R^2's obtained in these regressions.[34] We consider three different restricted models: one with no controls, another with only country fixed effects and the third containing both country and year effects. Results are reported in Table C1 in Appendix C. In all cases, the value of the ratio exceeded the cutoff value of 1, implying that selection bias is not likely to explain away our key results.

7.5. **Nonlinear Models.** Columns 4 and 5 in Table 6 replicate columns 4 and 7 in Table 1 using a logit specification.[35] The coefficients of the interactions of SIZE and the public and private payoffs maintain the expected signs and remain significant. In nonlinear specifications, however, one has to be cautious when interpreting the change in two interacted variables, as Ai and Norton (2003) pointed out. Appendix C.1 discusses this issue in more detail and shows that our conclusions still hold when nonlinear estimation is considered.

7.6. **Alliances in Conflict.** It may so happen that in some cases, *alliances* of groups could form. For instance, in the First Sudanese Civil War, also known as the Anyanya Rebellion, a conglomeration of the Acholi, Bari, Dinka, Lotuko, Madi, Nuer and the Zande from South Sudan came together, albeit in an alliance marked by substantial infighting. Other alliances are not hard to find: e.g., ethnic alliances exist in the Casamance conflict in Senegal or in the Liberian war that toppled the Taylor government.

As already described, the data we use code ethnic groups in conflict against the State. In the case of alliances, *each* ethnic group is so coded. As expected, the dataset has a number of such conflicts. Now, several of these conflicts are genuinely separate conflicts, and some are not. It is unclear how one might approach this problem comprehensively without running into severe issues of endogeneity in the definition of a "group."

Without pretending to satisfactorily solve this dilemma, one can run a rough variant of our exercise by mechanically combining all multiple instances of conflict. Table 7 replicates Table 1 using an alternative definition of group size, $\text{SIZE}_{\text{COAL}}$. This variable is defined as follows: for peace years, $\text{SIZE}_{\text{COAL}}$ and SIZE coincide. For years where *some* group is in conflict, $\text{SIZE}_{\text{COAL}}$ adds up the size of *all* the groups in conflict in the same country and year. In this way we try to capture the possibility that there exists an alliance between the fighting groups. The variable OIL_{COAL} is defined in a similar way: in peace years, OIL and OIL_{COAL} are identical. In case of conflict, $\text{SIZE}_{\text{COAL}}$ adds up the oil in the homelands of all the groups in conflict in that country and year. Our conclusions are robust to this variation.

7.7. **Further Variations.** As mentioned before, the Online Appendix (Appendix C) contains some additional tests. More specifically, we provide results clustering errors at the country level

[34]An additional R^2 is needed in the calculation, R_{max}, which is the maximum R^2 that could be attained if all the relevant controls were observed. We use a conservative value for this quantity and settle R_{max}=1.

[35]Given the difficulties involved in interpreting interactions in nonlinear models, for simplicity we focus on models containing one interaction only.

Dependent Variable: **Conflict Incidence**					
	[1]	[2]	[3]	[4]	[5]
$\text{SIZE}_{\text{COAL}}$	0.022***	0.041***	0.059***	0.064***	0.038***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
OIL_{COAL}	0.002***	0.003***	0.004***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\text{SIZE}_{\text{COAL}} \times \text{OIL}_{\text{COAL}}$		-0.005***	-0.005***	-0.005***	-0.004***
		(0.000)	(0.000)	(0.000)	(0.000)
$\text{SIZE}_{\text{COAL}} \times \text{AUTOC}$					0.058***
					(0.001)
GIP			-0.021***	-0.023***	-0.024***
			(0.000)	(0.000)	(0.000)
GROUPAREA			-0.006***	-0.008***	-0.007**
			(0.002)	(0.002)	(0.017)
SOILCONST			-0.001**	-0.001	-0.001*
			(0.049)	(0.272)	(0.056)
DISTCAP			2.684***	3.704***	3.200***
			(0.000)	(0.000)	(0.000)
MOUNT			0.003**	0.002	0.003*
			(0.038)	(0.145)	(0.085)
PARTITIONED			-0.000	-0.001	-0.001
			(0.685)	(0.447)	(0.438)
GDP				0.002*	0.003**
				(0.064)	(0.010)
POP				0.000	0.002
				(0.883)	(0.333)
LAG	0.886***	0.885***	0.877***	0.874***	0.876***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-0.027***	-0.039***	-0.061***	-0.052	-0.107**
	(0.000)	(0.000)	(0.000)	(0.273)	(0.013)
R^2	0.845	0.845	0.847	0.851	0.854
Obs	64839	64839	64839	53988	55383

Table 7. Group Size and Conflict: Coalitions. This table regresses conflict incidence on group size (allowing for the possibility of coalitions) and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain year dummies and country fixed effects. The coefficients of GROUPAREA and DISTCAP have been multiplied by 1000. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

and at the group and country level (two-way clustering) and also dropping particular world regions; see Tables C2, C3 in Appendix C. Our conclusions are in general robust to these variations. Finally, we also examine the potential confounding role of ethnic fractionalization and polarization; see Section C.5.

8. Conclusion

Group size matters in social conflict. But there is more than one view on just how it matters. In the introduction to his essay, "On Liberty," John Stuart Mill (1859) writes:

"Society . . . practices a social tyranny more formidable than many kinds of political oppression, since, though not usually upheld by such extreme penalties, it leaves fewer means of escape, penetrating much more deeply into the details of life, and enslaving the soul itself. Protection, therefore, against the tyranny of the magistrate is not enough; there needs protection also against the tyranny of the prevailing opinion and feeling, against the tendency of society to impose, by other means than civil penalties, its own ideas and practices as rules of conduct on those who dissent from them . . . ”

Mill is referring to the tyranny of the majority, a notion that also appears in the writings of John Adams and in the Federalist Papers, in the 18th century, and then amplified and used more extensively by Alexis de Tocqueville (1835).

Arrayed against this distinguished company are Wilfredo Pareto and Mancur Olson, who emphasize the power of minorities to cohere around a cause. In the words of Pareto (1927, p. 379), who was remarking on protectionist tendencies in trade,

“[A] protectionist measure provides large benefits to a small number of people, and causes a very great number of consumers a slight loss. This circumstance makes it easier to put a protection measure into practice.”

In this paper we’ve studied a model of social conflict with two main features: there are *multiple potential threats* to peace (i.e., a coalition might form around one or more characteristics), and the conflict may be over a *public* or a *private* good or a mixture of the two. Despite the fact that conflict is inefficient, we show that it may be an outcome even under complete information. The reason is that the existence of several conflictual divisions in society might make impossible to find an arrangement that simultaneously prevents all such threats to peace.

The multiplicity of threats to the established order, and the consequent inability of society to generate sustained peace, is an important theme of this paper. It is, however, not the theme of central empirical import. That has to do with the groups that stand at the forefront of hindrance to peace. The main empirical implication of the theory is that large groups are more likely to be involved in conflict against the State when the prize is public, while small groups are more likely to be involved in conflict against the State when the prize is private. By using a global panel dataset at the ethnic group level we find significant and powerful empirical support for these claims.

REFERENCES

Ai, C. R. and E. C. Norton (2003), "Interaction terms in logit and probit models," *Economics Letters* **80**: 123–129.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy* **80**:151–184.

Angrist, J. D., and J. Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, **24**: 3-30.

BALIGA, S. AND T. SJOSTROM (2004), "Arms Races and Negotiations, *Review of Economic Studies* **71**, 351–369.

Bellows, J., and E. Miguel (2009), "War and Local Collective Action in Sierra Leone". *Journal of Public Economics* **93**:1144?1157.

Berman N., M. Couttenier, D. Rohner and M. Thoenig (2015), "This mine is mine! How minerals fuel conflict in Africa," Mimeo.

Besley T. and T. Persson (2011), "The logic of political violence," *Quarterly Journal of Economics* **126**: 1411–1445.

Bester, H. and K. Wärneryd (2006), "Conflict and the Social Contract," *Scandinavian Journal of Economics* **108**, 231–249.

Blundell R. and S. Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics* **87**: 115–143.

Bondareva, O. N. (1963), "Some applications of linear programming methods to the theory of cooperative games", *Problemy Kybernetiki* **10**: 119–139.

Bruk, S., and V. S. Apenchenko (eds) (1964), *Atlas Narodov Mira [Atlas of the Peoples of the World]*. Moscow: Glavnoe Upravlenie Geodezii i Kartogra.

Cederman, L.-E., H. Buhaug and Rod, J. K. (2009), "Ethno-Nationalist Dyads: A GIS-Based Analysis," *Journal of Conflict Resolution* **53**: 496-525.

Chamberlin, J. (1974), "Provision of Collective Goods as a Function of Group Size," *American Political Science Review* **68**: 707–716.

Doyle, M. and N. Sambanis (2006), *Making War and Building Peace*. Princeton: Princeton University Press.

Dube, O. and J. Vargas (2013), "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *Review of Economic Studies* **80**, 1384–1421.

Esteban, J. and D. Ray (2008), "On the Salience of Ethnic Conflict," *American Economic Review* **98**: 2185–2202.

Esteban, J. and D. Ray (2001), “Collective Action and the Group Size Paradox,” *American Political Science Review* **95**: 663–672.

Esteban, J. and D. Ray (2011), “Linking Conflict to Inequality and Polarization,” *American Economic Review* **101**: 1345–74.

Esteban, J., Mayoral, L. and D. Ray (2012),“Ethnicity and Conflict: An Empirical Investigation,” *American Economic Review* **102**: 1310–1342.

Fearon, J. (1995), “Rationalist Explanations for War, *International Organization* **49**, 379–414.

Fearon, J. (2003), “Ethnic and cultural diversity by country,” *Journal of Economic Growth* **8**:195-222.

Fearon, J. (2005), “Primary Commodity Exports and Civil War,” *Journal of Conflict Resolution* **49**, 483–507

Fearon, J. and D. Laitin (2003), “Ethnicity, Insurgency, and Civil War,” *American Political Science Review* **97**: 75-90.

Fischer, G., F. Nachtergaele, S. Prieler, H. van Velthuizen, L. Verelst, and D. Wiberg (2008), *Global Agro-ecological Zones Assessment for Agriculture*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Freedom House (2014), *Freedom of the World*, dataset, `https://freedomhouse.org/report/freedom-world-aggregate-and-subcategory-scores`.

Garfinkel, M. and S. Skaperdas (2000), “Conflict Without Misperceptions or Incomplete Information: How the Future Matters,” *Journal of Conflict Resolution* .**44**, 793–807

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002), “Armed Conflict 1946-2001: A New Dataset,” *Journal of Peace Research* **39**: 615–637.

IntierraRMG (2015), “SNL Metals and Mining”, dataset,
`http://www.snl.com/Sectors/metalsmining/Default.aspx`.

Jackson, M. and M. Morelli (2007), “Political Bias and War,” *American Economic Review* **97**, 1353–1373.

Kirshner, J. (2000), “Rationalist Explanations for War?,” *Security Studies* **10**: 143–150.

Le Billon, P. (2001), “The Political Ecology of War: Natural Resources and Armed Conflicts,” *Political Geography* **20**, 561–584.

Licklider, R. (1995), “The Consequences of Negotiated Settlements in Civil Wars, 1945-1993,” *American Political Science Review* **89**: 681-690.

Lujala, P. (2010), “The Spoils of Nature: Armed Civil Conflict and Rebel Access to Natural Resources,” *Journal of Peace Research* **47**, 15–28.

Lujala, P., J. K.Rod and N. Thieme (2007), “Fighting over Oil: Introducing a New Dataset,” *Conflict Management and Peace Science* **24**: 239-56.

McGuire, M.C. (1974), “Group Size, Group Homogeneity and the Aggregate Provision of a Pure Public Good under Cournot Behavior,” *Public Choice* **18** (Summer): 107–126.

Marwell, G. and P. Oliver (1993), *The Critical Mass in Collective Action. A Micro-Social Theory*. Cambridge, UK: Cambridge University Press.

Mill, J. S. (1859), *On Liberty*, London: Parker and Son.

Morelli, M. and D. Rohner (2015), “Resource Concentration and Civil Wars,” *Journal of Development Economics* **117**: 32–47.

Oliver, P. and G. Marwell (1988), “The Paradox of Group Size in Collective Action: A Theory of the Critical Mass. II,” *American Sociological Review* **53**: 1–8.

Olson, M. (1965), *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Oster, E. (2016), “Unobservable Selection and Coefficient Stability: Theory and Evidence”. *Journal of Business and Economic Statistics*, forthcoming.

Pareto, V. [1906] (1927), *Manual of Political Economy*. New York: A.M. Kelley.

Polity IV (2014), “Political Regime Characteristics and Transitions”, dataset, `http://www.systemicpeace.org/inscrdata.html`.

Powell, R. (2004), “The Inefficient Use of Power: Costly Conflict with Complete Information,” *American Political Science Review* **98**, 231–241.

Powell, R. (2006), “War as a Commitment Problem,” *International Organization* **60**, 169–203.

Religion and State Project ARDA, Round 2, Dataset: `http://www.thearda.com/ras/`.

Sambanis, N. (2001), “Do Ethnic and Non-Ethnic Civil Wars Have the Same Causes? A Theoretical and Empirical Enquiry,” Journal of Conflict Resolution **45**: 259-282.

Sánchez-Pagés, S. (2009), “Conflict as Part of the Bargaining Process,” *Economic Journal* **119**, 1189–1207.

Sandler, T. (1992), *Collective Action: Theory and Applications*. Ann Arbor, MI: University of Michigan Press.

Scarf, H. (1967), “The Core of an n-Person Game,” *Econometrica* **35**, 50–69.

Schneider, G.and N. Wiesehomeier (2006), “Ethnic Polarization, Potential Conflict, and Civil Wars: Comment,” unpublished manuscript, University of Konstanz.

Shapley, Ll. S. (1967), “On Balanced Sets and Cores,” *Naval Research Logistics Quarterly* **14**: 453–460.

Skaperdas, S. (1996) “Contest Success Functions,” *Economic Theory* **7**: 283–290.

Slantchev, B. (2003), “The Power to Hurt: Costly Conflict with Completely Informed States,” *American Political Science Review* **47**, 123–135.

Taylor, M. (1987) *The Possibility of Cooperation.* Cambridge, UK: Cambridge University Press.

Tocqueville, A. de (1835), *Democracy in America*, Chicago, IL: University of Chicaco Press (2000 edition).

Vreeland, J. R. (2008). "The Effect of Political Regime on Civil War," *Journal of Conflict Resolution*, **52**: 401–425.

Weidmann, N. B., Rod, J.K. and Cederman, L. -E. (2010), "Representing Ethnic Groups in Space: A New Dataset," *Journal of Peace Research* **47**: 491-499.

Wooldridge, J.M. (2003), "Cluster-Sample Methods in applied Econometrics," *The American Economic Review* **93**, Papers and Proceedings of the One Hundred Fifteenth Annual Meeting of the American Economic Association, 133–138.

Wucherpfennig, J., Weidmann, N., Girardin, L., Cederman, L. and A. Wimmer (2011), "Politically relevant ethnic groups across space and time: Introducing the GeoEPR dataset," *Conflict Management and Peace Science* **28**: 423–437.

APPENDIX A.

Proof of Proposition 2. Recall that groups are indexed in decreasing order of size, so that $m_i \geq m_{i+1}$ for all i. Let $\sigma_i \equiv m_i/(1-m_i)$. Also, for each $i \geq 2$, define

$$\gamma_i \equiv \left(\frac{1-m_i}{m_1}\right)^{1/\alpha}$$

and let $\gamma_1 \equiv [(1-m_1)/m_2]^{1/\alpha}$.[36] Our proof will rely on

Lemma 1. *Assume that the prize is public. Then Rebel i challenges an uniform allocation to it (that is, a total of $m_i \Psi m_1$) if and only if*

$$\frac{\sigma_i\gamma_i(\sigma_i\gamma_i + k)}{(\sigma_i\gamma_i+1)^2} > m_1. \tag{17}$$

Moreover, if condition (17) holds for any i, then it also holds for $j < i$. That is, if any Rebel wishes to initiate conflict, so does a larger Rebel, provided both receive their shares of the uniform allocation.

Proof. Suppose that Rebel i initiates a conflict. If $i \geq 2$, then m_1 is the size of the largest group that intersects the defendant, and writing $\mu_i \equiv m_1/(1-m_i)$, we know that the per-capita prizes are given by Ψ and $\Psi\mu_i$ to Rebel and State respectively. Using (3),

$$\frac{r}{\bar{r}} = \mu_i^{-1/\alpha} = \gamma_i, \tag{18}$$

and then using (5), the win probability for the Rebel is given by

$$p = \frac{m_i\gamma_i}{m_i\gamma_i + (1-m_i)}.$$

Overall expected payoffs per-capita to Rebel i are therefore given by

$$\Psi\left[k\frac{m_i\gamma_i}{m_i\gamma_i + (1-m_i)} + (1-k)\left(\frac{m_i\gamma_i}{m_i\gamma_i + (1-m_i)}\right)^2\right], \tag{19}$$

and under the assumption that the no-conflict allocation to the Rebel is uniform, we see that conflict will occur if and only if

$$k\frac{m_i\gamma_i}{m_i\gamma_i + (1-m_i)} + (1-k)\left(\frac{m_i\gamma_i}{m_i\gamma_i + (1-m_i)}\right)^2 > m_1. \tag{20}$$

Dividing through by $1-m_i$ above and below on the left-hand side of this inequality and using the definition $\sigma_i = m_i/(1-m_i)$, elementary manipulation shows that (20) is equivalent to (17). This completes the proof of the first part of the lemma.

Let $\eta_i \equiv \sigma_i\gamma_i$; then (17) is equivalent to the condition

$$m_1 < \frac{\eta_i(\eta_i+k)}{(\eta_i+1)^2},$$

[36] The reason for this asymmetry stems from the fact that the largest group intersecting the State is of size m_1 for all Rebels $i \neq 1$, whereas for $i = 1$ it is of size m_2.

and it is easy to see that the right-hand side of this inequality is increasing in η_i.[37] So to complete the proof, it suffices to show that $\eta_i \geq \eta_{i+1}$ for all $i \geq 1$. For $i \geq 2$,

$$\eta_i = \sigma_i \gamma_i = \frac{m_i}{m_1^{1/\alpha}(1 - m_i)^{(\alpha-1)/\alpha}} \geq \frac{m_{i+1}}{m_1^{1/\alpha}(1 - m_{i+1})^{(\alpha-1)/\alpha}} = \eta_{i+1},$$

while

$$\eta_1 = \sigma_1 \gamma_1 = \frac{m_1}{m_2^{1/\alpha}(1 - m_1)^{(\alpha-1)/\alpha}} \geq \frac{m_2}{m_1^{1/\alpha}(1 - m_2)^{(\alpha-1)/\alpha}} = \eta_2,$$

where the inequality above is a matter of simple algebra. ■

Lemma 1 implies that a necessary and sufficient condition for conflict-proneness is the inequality (17) applied to $i = 1$, so that the largest group wishes to enter into conflict under a uniform allocation. Setting $i = 1$ in (17) and manipulating, we obtain (8), as required. ■

Proof of Proposition 5. Let $\mathcal{C}$ be a balanced collection such that (13) holds for the smallest group in it. Then it is easy to see that (17) holds for $i = s$, and by Lemma 1, (17) holds for *every* group in the collection. So each group of size m will challenge any allocation to it that has no higher aggregate value than $m\Psi m_1$. To prevent conflict, then, there must exist some feasible allocation $\mathbf{x}$ of Ψm_1 such that for every group $G \in \mathcal{C}$,

(21)
$$\int_{j \in G} x(j) > m\Psi m_1.$$

The remainder of the argument mimics the proof of Proposition 4, where we set v in that proof equal to Ψm_1 here. That argument tells us that

$$\int_j x(j) > \Psi m_1,$$

a contradiction to the feasibility of $\mathbf{x}$. ■

[37]Consider the function $x(x + k)/(x + 1)^2$. Differentiate with respect to x to see that the derivative is positive whenever $k \in [0, 1]$.

APPENDIX B.

This Appendix contains detailed definitions of all the variables employed in the empirical analysis (Section B.1) as well as a table of summary statistics (Section B.2).

B.1. **Variable Definitions.** Conflict *onset*: group-level dummy variable that equals 1 in a given year if an armed conflict against the state resulting in more than 25 battle-related deaths involving that ethnic group newly starts. For ongoing wars, ONSET is coded as missing. Source: CBR.

Conflict *incidence*: group-level dummy variable equal to 1 for those years where an ethnic group is involved in armed conflict against the state resulting in more than 25 battle-related deaths. Source: CBR.

SIZE: relative size of the group, source: CBR.

OIL: log of the homeland area covered by oil (in thousands of square kilometres) times the international price of oil. To avoid taking the log of zero, 1 has been added to all observations. Source: information on oil fields comes from PETRODATA (Lujala et al. 2007). Data on oil prices comes from the World Bank.

OIL(AREA): log of the homeland area covered by oil (in thousands of square kilometres). To avoid taking the log of zero, 1 has been added to all observations. Source: PETRODATA (Lujala et al. 2007).

OIL(SHARE): ratio of OIL(AREA) and the total area of the homeland. Source: PETRODATA and GREG.

MINES: measures mineral availability in the ethnic homeland and is computed in the following way. First, we consider 13 minerals (bauxite, coal, copper, diamond, gold, iron, lead, nickel, platinum, phosphate, silver, tin and zinc) for which international price data is readily available. For each mineral, year and ethnic group, we create a dummy variable that is one if the group has at least one active mine of that mineral. Then, each of these dummies is multiplied by its normalized international price. The latter is constructed as the log of its international price divided by the log of its price in 1980 (the year when the data starts). The variable MINES is computed as the sum of the resulting products. Data on mineral availability comes from the *Raw Material Data* dataset, (IntierraRMG, 2015) whereas data on mineral prices is provided by the World Bank.

MINES+OIL: constructed in the same way as MINES but it includes 14 minerals and energy products, being oil one of them. Source: PETRODATA, Raw Material Data and World Bank.

MINES(UNWEIGH.): analogous to MINES, except that it doesn't include information on prices. That is, it is created as the sum of the 13 mineral dummies (not weighted by prices).

MINES+OIL(UNWEIGH.): similar to MINES(UNWEIGH.), but it also includes a dummy for oil availability.

AUTOC: country average of the Polity IV autocracy index for the years 1945 to 1960. For countries that didn't exist before 1960, the first non-missing autocracy observation is used instead.

AUTOC$_{1945-60}$: country average of the Polity IV autocracy index for the years 1960 to 1980.

EXCLUDED: dummy variable that is 1 if the ethnic group is in power in a given country and year (lagged one year), source: CBR.

EXCLUDED$_{1945-60}$: average of EXCLUDED for the years 1945 to 1960. For countries that didn't exist before 1960, the first non-missing value value has been used instead.

EXCLUDED$_{1960-80}$: average of EXCLUDED for the years 1960 to 1980.

PUB(EMR): PUB index from Esteban et al. (2012).

RELIGFREEDOM: this variable measures the extent to which, in practice, a state is willing to restrict some or all religions. It is measured on a 1–6 scale and high values reflect a higher degree of religious freedom. Source:*Religion and State project*, ARDA (`http://www.thearda.com/ras/`).

GIP: dummy variable that is 1 if the ethnic group is in power in a given country and year (lagged one year), source: CBR.

GROUPAREA: area of the ethnic homeland (in thousands of square kilometres), source: GREG.

AREA(SHARE): area of the ethnic homeland relative to total area of the country, source: GREG.

SOILCONST: a measure of the limitations that the group's soil presents to agriculture. It's constructed using the Harmonized World Soil Database from Fischer et al., (2008). Fisher et al. (2008) construct a global grid of 38 nutrient availability ranked from 1 (no or slight constraints) to 4 (very severe constraints), and also including categories 5 (mainly non-soil), 6 (permafrost area) and 7 (water bodies). SOILCONST is constructed as the average of the cell values pertaining to the group's homeland.

DISTCAP: group's distance to the country capital, source: CBR.

MOUNT: 0-1 index capturing the group's share of mountainous terrain, source: CBR.

PEACEYEARS: number of years since the last group-level onset and LAG is lagged conflict incidence, source:CBR.

PARTITIONED: dummy variable that is 1 if the ethnic homeland covers two or more countries, source: GREG.

GDP: log of (country-level) GDP per capita, lagged one year. Source: Penn World Tables.

POP: log of total country population (POP), lagged one year. Source: Penn World Tables.

B.2. **Summary Statistics.** Table B1 provides summary statistics.

	Obs	**Mean**	**SD**	**Min**	**Max**
incidence	64001	0.04	0.19	0.00	1.00
onset	61928	0.00	0.06	0.00	1.00
SIZE	64001	0.10	0.23	0.00	1.00
OIL	64001	1.06	2.13	0.00	12.38
OIL(AREA)	64001	0.31	0.82	0.00	6.95
OIL(SHARE)	64001	0.03	0.09	0.00	0.92
MINES	39670	0.95	1.72	0.00	13.00
MINES+OIL	39670	1.18	1.87	0.00	14.00
MINES(UNWEIGH.)	39670	1.02	1.85	0.00	13.00
MINES+OIL(UNWEIGH.)	39670	1.33	2.00	0.00	14.00
AUTOC	63890	0.43	0.33	0.00	1.00
$\text{AUTOC}_{1945-60}$	39218	0.54	0.29	0.00	1.00
EXCLUDED	63789	0.86	0.35	0.00	1.00
$\text{EXCLUDED}_{1945-60}$	64001	0.87	0.34	0.00	1.00
$\text{EXCLUDED}_{1960-80}$	38670	0.86	0.34	0.00	1.00
RELIGFREEDOM	25280	0.67	0.24	0.17	1.00
GIP	64001	0.14	0.35	0.00	1.00
GROUPAREA	64001	84.3	407	0.04	7355
AREA(SHARE)	61968	0.09	0.20	0.00	1.00
SOILCONST	64001	1.62	0.78	0.00	6.15
DISTCAP	64001	917	1030	5.00	7971
MOUNT	64001	0.37	0.36	0.00	1.00
PARTITIONED	64001	0.62	0.48	0.00	1.00
GDP	56945	7.75	1.16	5.08	11.16
POP	61893	17.08	1.81	11.73	20.98

Table B1. Summary Statistics

APPENDIX C. (ONLINE APPENDIX)

C.1. **Interactions in Nonlinear Models.** Interpreting the coefficients associated with interactions is straightforward in linear models, as they are simply the appropriate cross-partial derivatives of the dependent variable with respect to the relevant variables in the interaction. However, this logic does not extend to nonlinear models, as shown by Ai and Norton (2003). In non-linear models, the cross-partial derivative does not admit a simple interpretation, and important differences arise with respect to the linear case. First, the "true sign" of the interaction does not need to equal the sign of the cross-partial derivative. Second, the significance of that interaction cannot be tested with a simple t-test on the coefficient of the interaction term (in the regression). Third, given the nonlinearity, the value of the interaction term depends on all the independent variables of the model. See Ai and Norton(2003) for a discussion.

To overcome these difficulties and in order to facilitate the interpretation of the interactions reported in Table 6 in the main text, we have evaluated the cross-partial derivative at each of the points in our sample. Figure 5 plots the derivative of the dependent variable with respect to SIZE and OIL, using the specification in Column 4, Table 6. This figure shows that the cross-derivative is negative for most observations in our sample, a result that mimics the one obtained for the linear case. Figure 6 plots the z-statistics associated with the cross-partial derivative for each of the points in the sample, together with confidence bands (at the 90 per cent level). This figure shows that the effect is significant in most cases. Similar results are found when interpreting the interaction of SIZE and AUTOC. In this case, the cross-partial derivative is positive and significant for most of the observations.[38]

C.2. **Assessing the Importance of the Omitted Variable Bias.** As explained in the main text, we have applied Oster's (2016) technique to assess the importance of the omitted variable bias. Table C1 presents our results. The full model corresponds to our baseline specification (column 8 in Table 1). Restricted models I, II and III correspond to models with no controls, with country fixed effects and country and year fixed effects, respectively. The figures in the table are the values of the ratio of correlations, see Oster (2016) for a precise definition. Since all ratios are larger than 1, the standard cutoff, one can conclude that it is unlikely that our results are due to selection on unobservables.

C.3. **Alternative ways of clustering the error term.** Table C2 replicates Table 1 in the main text with clustering errors in two different ways. In columns 1 to 5, errors are clustered at the country level, while in columns 6-10 errors are clustered at the group (as opposed to country-group) and at the country level (two-way clustering). Notice that since the ethnic homeland is often split by an international border, the latter dimensions are not nested. Table C2 shows that our conclusions remain unchanged when other clustering schemes are considered.

[38] For the sake of brevity, we don't report the corresponding graphs as they are very similar to those associated with SIZE and OIL, but they are available upon request.

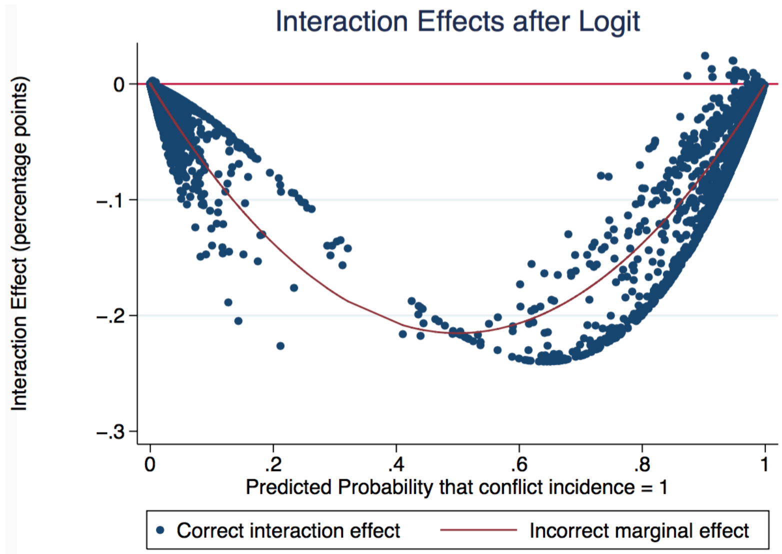


Figure 5. Interpreting Interactions in Nonlinear Models: Cross-Partials
This graph depicts the value of the cross-partial derivative of conflict incidence with respect to OIL and SIZE, for each of the points in the sample. Estimates from Table 6 (column 4) have been employed to compute the estimates.

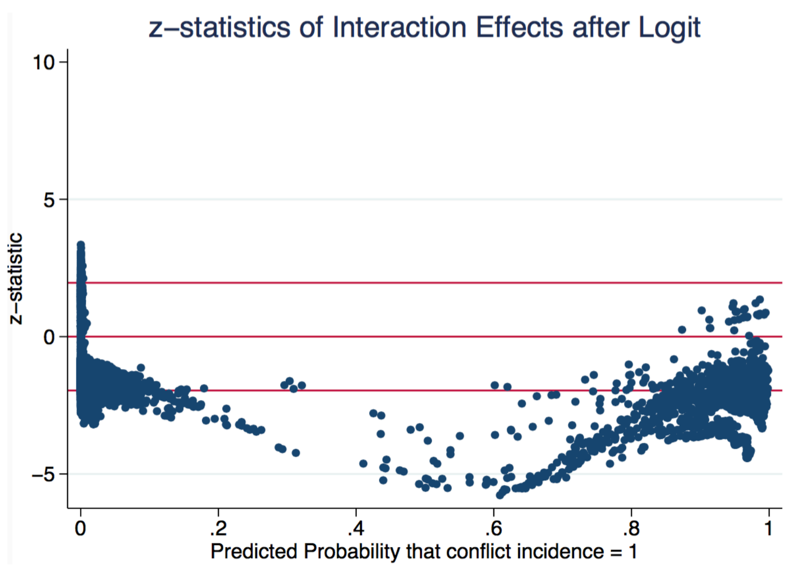


Figure 6. Interpreting Interactions in Nonlinear Models: z-Statistics
This graph plots the values of the z-statistics associated with each of the points of the cross-partial derivative reported in Figure 5.

C.4. **Dropping regions of the world.** Table C3 drops observations from particular regions of the world. Those regions are: former USSR countries (columns 1 and 2), Asia (columns 3 and 4), Middle East (columns 5 and 6), Sub-Saharan Africa (columns 7 and 8) and Latin America (columns 9 and 10). For each region, the first (second) column replicates column 4 (5) in Table

	Selection on Observed and Unobserved variables	
	SIZE× OIL	SIZE× AUTOC
RESTRICTED MODEL I	4.49754	-3.96099
RESTRICTED MODEL II	-19.56435	1.45680
RESTRICTED MODEL II	20.30606	1.43471

Table C1. Using Selection on Observables to Assess the Bias from Unobservables.
Notes. This table applies Oster's (2015) technique to assess how strong the correlation between the unobservables and oue key variables has to be in order to explain away our observed results. Calculations have been performed using the software *psacalc* provided by the author. See Oster (2015) for details.

	Dependent Variable: **Conflict Incidence**									
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
SIZE	-0.002	0.003	0.007***	0.007***	0.003	-0.002	0.003	0.007***	0.007***	0.003
	(0.366)	(0.194)	(0.003)	(0.004)	(0.156)	(0.372)	(0.192)	(0.003)	(0.004)	(0.215)
OIL	0.448	0.684*	0.830**	0.795*	0.762**	0.448	0.684*	0.830**	0.795*	0.762*
	(0.139)	(0.078)	(0.027)	(0.061)	(0.010)	(0.143)	(0.082)	(0.030)	(0.066)	(0.077)
SIZE× OIL		-1.363**	-1.528***	-1.521**	-1.390***		-1.363**	-1.528***	-1.521**	-1.390**
		(0.021)	(0.007)	(0.014)	(0.000)		(0.023)	(0.007)	(0.015)	(0.022)
SIZE× AUTOC					0.009**					0.009**
					(0.015)					(0.037)
GIP			-0.003*	-0.003	-0.003*			-0.003*	-0.003	-0.003
			(0.087)	(0.128)	(0.057)			(0.081)	(0.120)	(0.118)
GROUPAREA			0.000	0.000	0.000			0.000	0.000	0.000
			(0.469)	(0.277)	(0.219)			(0.468)	(0.270)	(0.303)
SOILCONST			-0.001	-0.000	-0.000			-0.001	-0.000	-0.000
			(0.240)	(0.640)	(0.472)			(0.241)	(0.642)	(0.594)
DISTCAP			0.001	0.002*	0.002***			0.001	0.002*	0.002
			(0.143)	(0.097)	(0.000)			(0.142)	(0.097)	(0.108)
MOUNT			0.002**	0.002**	0.002			0.002**	0.002**	0.002*
			(0.015)	(0.035)	(0.130)			(0.016)	(0.036)	(0.052)
PARTITIONED			-0.001	-0.001	-0.001			-0.001	-0.001	-0.001
			(0.540)	(0.279)	(0.243)			(0.542)	(0.281)	(0.245)
GDP				0.001	0.003***				0.001	0.003**
				(0.393)	(0.006)				(0.391)	(0.035)
POP				0.001	0.001				0.001	0.001
				(0.793)	(0.710)				(0.792)	(0.868)
LAG	0.895***	0.895***	0.894***	0.893***	0.898***	0.895***	0.895***	0.894***	0.893***	0.898***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-0.002	-0.005**	-0.009***	-0.034	-0.041	-0.002	-0.005**	-0.009***	-0.034	-0.041
	(0.355)	(0.023)	(0.000)	(0.695)	(0.319)	(0.355)	(0.021)	(0.000)	(0.694)	(0.626)
R^2	0.844	0.844	0.844	0.846	0.851	0.844	0.844	0.844	0.846	0.851
Obs	64839	64839	64839	57559	55383	64839	64839	64839	57559	55383

Table C2. Group Size and Conflict: Baseline.
Notes. This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain year dummies and country fixed effects. Robust standard errors clustered at the country level have been computed in columns 1–5. Columns 6 to 10 use two-way clustering with standard errors clustered at the group *and* the country level. p-values are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

1, that is we consider specifications with and without the interaction of group size and the public prize. Results are general robust, except than dropping countries in Sub-Saharan Africa reduces the significance of the interaction of size and oil.

	Dependent Variable: **Conflict Incidence**									
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
SIZE	0.007***	0.003	0.007***	0.004*	0.007***	0.003	0.003	0.001	0.010***	0.005**
	(0.001)	(0.147)	(0.002)	(0.068)	(0.001)	(0.147)	(0.187)	(0.639)	(0.000)	(0.050)
OIL	0.837***	0.801**	1.014***	0.961***	0.837***	0.801**	0.383	0.362	1.077***	1.029***
	(0.008)	(0.011)	(0.007)	(0.008)	(0.008)	(0.011)	(0.125)	(0.156)	(0.004)	(0.007)
SIZE× OIL	-1.623***	-1.438***	-1.758***	-1.646***	-1.623***	-1.438***	-0.296	-0.268	-2.231***	-2.051***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.381)	(0.432)	(0.000)	(0.000)
SIZE× AUTOC		0.010**		0.009*		0.010**		0.006**		0.010**
		(0.022)		(0.072)		(0.022)		(0.022)		(0.026)
GDP	0.002*	0.003***	-0.000	0.002	0.002*	0.003***	0.002**	0.002**	0.001	0.003**
	(0.065)	(0.002)	(0.903)	(0.195)	(0.065)	(0.002)	(0.013)	(0.014)	(0.172)	(0.019)
POP	0.003	0.002	0.000	0.000	0.003	0.002	-0.009***	-0.009***	0.005**	0.005**
	(0.220)	(0.313)	(0.850)	(0.950)	(0.220)	(0.313)	(0.000)	(0.000)	(0.029)	(0.044)
legip	-0.003*	-0.003*	-0.002	-0.002	-0.003*	-0.003*	-0.007***	-0.007***	-0.003	-0.003
	(0.061)	(0.061)	(0.316)	(0.277)	(0.061)	(0.061)	(0.006)	(0.006)	(0.113)	(0.120)
GROUPAREA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.135)	(0.153)	(0.415)	(0.295)	(0.135)	(0.153)	(0.208)	(0.183)	(0.367)	(0.380)
SOILCONST	-0.000	-0.000	0.001	0.001	-0.000	-0.000	-0.001***	-0.001***	-0.000	-0.000
	(0.825)	(0.749)	(0.250)	(0.305)	(0.825)	(0.749)	(0.001)	(0.001)	(0.436)	(0.389)
DISTCAP	0.002***	0.002***	0.000	0.000	0.002***	0.002***	0.002***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.333)	(0.408)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
MOUNT	0.002	0.002	0.001	0.001	0.002	0.002	0.002*	0.002*	0.002	0.002
	(0.123)	(0.138)	(0.353)	(0.451)	(0.123)	(0.138)	(0.097)	(0.092)	(0.119)	(0.143)
PARTITIONED	-0.001	-0.001	-0.002	-0.002	-0.001	-0.001	0.001	0.001	-0.002	-0.002
	(0.364)	(0.321)	(0.199)	(0.154)	(0.364)	(0.321)	(0.290)	(0.290)	(0.186)	(0.161)
LAG	0.896***	0.900***	0.884***	0.890***	0.896***	0.900***	0.897***	0.897***	0.894***	0.899***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-0.060*	-0.039	-0.010	-0.035	-0.060*	-0.039	0.166***	0.074***	-0.081**	-0.099***
	(0.077)	(0.423)	(0.824)	(0.445)	(0.077)	(0.423)	(0.000)	(0.006)	(0.018)	(0.005)
R^2	0.849	0.854	0.849	0.856	0.849	0.854	0.862	0.863	0.828	0.834
Obs	55261	53085	39689	38011	55261	53085	38532	37819	46815	44767

Table C3. Dropping regions of the world
Notes. This tables reproduces columns 4 and 5 from Table 1 dropping regions of the world. Regions dropped are: former USSR countries (columns 1 and 2), Asia (columns 3 and 4), the Middle East (columns 5 and 6), Sub-Saharan Africa (columns 7 and 8) and Latin America (columns 9 and 10). All regressions contain year dummies and country fixed effects. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

C.5. **The Role of Ethnic Fractionalization and Polarization.** Esteban et al (2012) showed that indices of ethnic fractionalization and polarization are good predictors of conflict. It is therefore possible that the presence of oil in the homeland of a group is more conducive to conflict when the country is ethnically diverse. Because group size and fractionalization are likely to be negatively correlated, it is possible that the negative effect of the coefficient of the interaction of oil and size is merely picking up the fractionalization effect. In order to examine this possibility, Table C4 adds to our baseline specification (column 5 in Table 1) the interaction of OIL and FRAC, an index of fractionalization (column 1), and the interaction of OIL and POL, a polarization index (column 2).[39] Columns 3 and 4 use the international price of oil OILPRICE rather than OIL to compute these interactions. Our results remain robust to these variations.

	Dependent Variable: **Conflict Incidence**			
	[1]	[2]	[3]	[4]
SIZE	0.003	0.002	0.002	0.002
	(0.341)	(0.386)	(0.398)	(0.398)
SIZE× OIL	-1.165**	-1.518***	-1.507***	-1.533***
	(0.013)	(0.001)	(0.001)	(0.001)
SIZE×AUTOC	0.014**	0.013**	0.013**	0.013**
	(0.011)	(0.013)	(0.012)	(0.013)
OIL×FRAC	3.407***			
	(0.001)			
OIL×POL		0.656		
		(0.891)		
OILPRICE×FRAC			0.034*	
			(0.076)	
OILPRICE×POL				-0.345**
				(0.025)
OIL	-0.768	0.883**	0.925***	0.953***
	(0.124)	(0.026)	(0.005)	(0.004)
GDP	0.002**	0.002*	0.002**	0.002*
	(0.047)	(0.063)	(0.042)	(0.077)
POP	0.003	0.004	0.003	0.005*
	(0.255)	(0.173)	(0.264)	(0.096)
GIP	-0.004*	-0.003*	-0.003*	-0.003*
	(0.066)	(0.096)	(0.097)	(0.097)
GROUPAREA	0.000	0.000	0.000	0.000
	(0.114)	(0.322)	(0.341)	(0.355)
SOILCONST	-0.000	-0.000	-0.000	-0.000
	(0.651)	(0.524)	(0.516)	(0.507)
DISTCAP	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
MOUNT	0.002	0.002*	0.002*	0.002*
	(0.176)	(0.098)	(0.096)	(0.094)
PARTITIONED	-0.002	-0.002	-0.002	-0.002
	(0.152)	(0.101)	(0.101)	(0.100)
LAG	0.896***	0.897***	0.897***	0.897***
	(0.000)	(0.000)	(0.000)	(0.000)
OILPRICE			-0.000	0.000
			(0.945)	(0.193)
c	-0.087*	-0.090*	-0.083*	-0.097**
	(0.067)	(0.058)	(0.056)	(0.037)
R^2	0.853	0.853	0.853	0.853
Obs	45210	45210	45210	45210
R^2	0.853	0.853	0.853	0.853
Obs	45210	45210	45210	45210

Table C4. Interactions of the Private Prize, Fractionalization and Polarization
Notes. This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain year dummies and country fixed effects. FRAC and POL are country-level indices of ethnic fractionalization and polarization, respectively. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

[39] Fearon's (2003) ethnic group classification and group sizes have been employed to compute FRAC and POL, see Esteban et al (2012) for a description of these indices. Linguistic distances are used to compute POL.