GROUPS IN CONFLICT:

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

LAURA MAYORAL
Instituto de Análisis Económico, Barcelona

DEBRAJ RAY
New York University and University of Warwick

January 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. We show that groups that initiate conflict are likely to be small when the prize is private, and large when the prize is public. To examine these predictions empirically we construct a global dataset at the ethnic group level and focus on conflict that occurs along ethnic lines. By building indices of prize publicness and privateness along the lines developed in Esteban, Mayoral and Ray (2012), we show that 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 belligerent group might emerge to challenge the existing state of affairs. We ask two questions:

1. When might conflict be an outcome, even if it is costly and inefficient, and even though there is flexibility in the distribution of peacetime allocations?

2. Who threatens the peace? That is, are large groups or small groups more likely to destabilize a peaceful allocation by threatening conflict?

In the second part of the paper, we examine if our theoretical predictions regarding group size and conflict are supported by the data.

The central difficulty to peace that we emphasize is the presence 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 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.[2] For instance, a

[1]We are grateful to Joan Esteban for helpful comments on an earlier draft. Ray thanks the National Science Foundation for research support under grant number SES-1261560.

[2]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 (2001), Baliga and Sjöstrom (2004), Bester and Warneryd (2006) and Sánchez-Pagés (2009) all employ some version of a hidden-type model to obtain

society can set up institutions that can adequately deal with the question of class conflict, only to be confronted by threats from a religious or geographical subgroup.[3]

It is, of course, entirely possible that these different markers all delineate essentially the same division of people: for instance, "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.[4] Then a system of transfers set up to deal with one sort of division may be entirely useless when confronted with another.

Our second question asks which groups are the most potent in generating and maintaining conflict. This is, of course, a question that cannot be answered in full generality, as the identity and cohesion of various potential groupings represent deep questions that can only be resolved through empirical investigation. But there is one aspect of a group that commands special attention, and that can be examined both theoretically and empirically: group *size*. Must large groups or small groups be appeased in the Coasian quest for a universally peaceful allocation? 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 of those preferences.

In contrast, the Pareto-Olson thesis (see Pareto 1927 and Olson 1965) argues that small groups may be more effective than large groups: after all, the smaller the group, the larger the potential *per-capita* payoff. This sort of argument fundamentally depends on the presumption that the prize is an excludable private good, so that group size erodes per-capita payoffs. This point has been noted in various settings by Chamberlin (1974), McGuire (1974), Marwell and Oliver (1993), Oliver and Marwell (1988), Sandler (1992), Taylor (1987) and Esteban and Ray (2001).

In this paper, we describe a general setting that allows for both "tyrannies," and ask whether the threat of conflict is more likely to be initiated by a large or a small group. In deciding whether or not to issue that threat, a group must trade conflict payoffs against peace-time outcomes, comparing the two sets of payoffs. We do this for both conflicts over public and private goods. The resulting theoretical exercise has a sharp prediction: 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.

---

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.

[3]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.

[4]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.

In the second part of the paper, we test this prediction. 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 1952 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. We closely follow the approach taken in Esteban et al. (2012) to obtain proxies for private and public payoffs.

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. To measure publicness, we construct a country-level index based on the degree of power afforded to those who run the country. The main idea is that the more democratic the country is, the less power the government has, and consequently the lower valuation of the public payoff to conflict. We also employ data on religious freedom to construct an alternative index of publicness.

Our empirical 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. On the other hand, larger ethnic groups are more likely to participate in conflict when the (country-level) valuation of the public payoff is high. 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.

Of course, it is well known in the 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. These twin interactions — indeed, neither interaction — have been unexplored in the literature. They combine and reconcile the tyranny of the majority with the Pareto-Olson thesis.

In what follows, Section 2 introduces a baseline model of peace and conflict. Sections 3 and 4 analyze the relation between group size and conflict when conflict is over private or public goods, respectively. 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.

2. A Baseline Model of Peace and Conflict

2.1. **Peaceful Allocations.** A society is made up of a unit mass of individuals. It has a total value (or "budget"); call it v. This value may represent material or economic resources such as revenues from oil, or the value of ideological positions such as "a Hindu state." We will view v as the total "appropriable resources" of society. (There may be other non-appropriable human or physical resources which we normalize to zero for everyone.) The value may be transferable to different degrees; we will return to this point below. For now let V denote the space of all feasible (peaceful) allocations of the form $\mathbf{x} = \{x(i)\}$, where i indexes a person. It is a subset of $\{\mathbf{x} | \int x(i) \leq v\}$. These represent allowable distributions of the aggregate value among individual members. An allocation is *unbiased* if every individual receives the same amount under it. In most of the cases that we consider, an unbiased allocation will be available.

2.2. **Conflict.** There is a finite collection of groups, each a *strict* subset of society and demarcated by ethnicity, geography, caste, religion, occupation and so on. Everyone belongs to at least one group. Each group can precipitate costly conflict to seize appropriable resources. In that case, we suppose that society is partitioned into two subsets, one of size m (pertaining to the "initiator" group) and the remainder of size $\overline{m}$ ($m + \overline{m} = 1$), and that they engage in a bilateral conflict. One might interpret this initiator as a group waging conflict against society as a whole, with the "defender" — the other group — being a proxy for the incumbent government. Or the initiator may be a majority group threatening violence against the defender, a minority.

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 a group seizes the aggregate value v, it can distribute that value among its members in an entirely costless fashion. Therefore, it is assumed that a group leader on each side extracts these resources from everyone to maximize per-capita group payoff.[5] Because the cost of effort provision is strictly convex, the group leader will ask for equal effort from each individual, and will make transfers if needed to compensate them.

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

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

[5] It is easy to write down variants 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).

where r is contribution per person in the initiating group, 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 defender.)

As already discussed, the initiator seeks to maximize its per-capita payoff

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

where π is the per-capita value of the prize to the initiator. A similar problem is faced by the defender, with a per-capita prize of $\overline{\pi}$.[6] 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 initiator, and by

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

for the defender. Conditions (1) and (2) yield a simple expression for the provision of individual resources by the initiator, 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 initiator, 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 initiator's payoff in conflict equilibrium. A parallel expression holds for the defender.

Conflict is a threat to peace, and we seek conditions under which that threat might manifest itself. Say that a peaceful allocation $\mathbf{x} \in V$ is *blocked* if there exists an initiator G and a corresponding conflict situation as described above such that the expected payoff to the group exceeds what it receives under the peaceful allocation:

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

A society is *prone to conflict* if the unbiased peaceful allocation is blocked. It is *actively conflictual* if every peaceful allocation, unbiased or not, is blocked.

---

[6] 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.

A reader schooled in cooperative game theory will immediately see this as analogous to the problem of whether the core of a suitably defined partition-function game is empty or not, and indeed it is. There is nothing here that is conceptually new, except that we build such a partition function from a precisely stipulated noncooperative game of conflict.

This is the framework we employ for addressing two questions. First, is conflict inevitable? We ask this question in a "weak" and "strong" sense. The weak sense is one in which we presume that the society is committed to equal treatment for all. If equal treatment is blocked, we say that our society is prone to conflict. The strong sense comes from dropping equal treatment. Indeed, because conflict is costly, for each conflictual outcome there is a peaceful outcome that Pareto-dominates it, provided that transfers are available. But is there one outcome that can *simultaneously* withstand all threats? This is the issue of whether the society is actively conflictual, as defined above. The fact 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.

Our second, equally important focus lies in understanding whether small or large groups initiate conflict. To do so, we must compare the payoffs of peaceful 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, with some caveats and qualifications that we discuss in Section 5.1.

3. Group Size and Conflict: Private Goods

3.1. **Unbiased Peaceful Allocations and Conflict-Proneness.** Little by way of additional interpretation is needed when the entire prize v is a private good; say, oil revenues. In this case we 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 unbiased allocation gives precisely v to each individual.

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

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

Using this information in (3), we see that

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

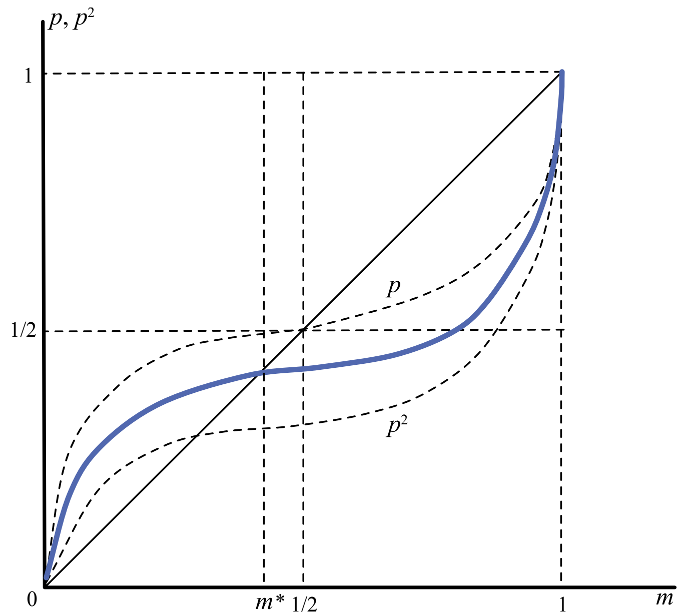


Figure 1. The Threshold m^* in Proposition 1

so that by (5),

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

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

Notice from (6) that smaller groups 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,

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

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 unbiased, 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 group 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.[7] 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,

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

for *any* $m \geq 1/2$.[8] This argument, in conjunction with Figure 1, shows that there is a unique intersection (crossing from above to below) in the interior of $(0, 1/2)$.[9] The proof of the proposition is now complete. ■

3.2. **Arbitrary Peaceful Allocations and Actively Conflictual Societies.** Proposition 1 shows that when the peacetime allocation is unbiased, the presence of small minorities makes the society prone to conflict. Yet an unbiased 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 peaceful allocation, because conflict is costly and therefore inefficient. (That allocation may or may not be unbiased.) A key question, therefore, is whether there is some allocation that *simultaneously* avoids conflict from *all* potential initiators. If the variety of potential threats is large relative to the degree of inefficiency, the answer is in the negative. In what follows we attempt to make this claim precise.

To formalize the idea of a "variety of threats", say that a finite collection $\mathcal{C}$ of potential initiators is *balanced* if there is a set of weights in $[0, 1]$, $\{\lambda(G)\}_{G \in \mathcal{C}}$, such that

$$\sum_{G \in \mathcal{C}, i \in G} \lambda(G) = 1 \text{ for every } i \text{ in society.} \tag{8}$$

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 groups 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, (8) 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:

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

[7] 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 $n = 1$.

[8] 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.

[9] 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.

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 (9), it must be the case that

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

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 we partition $[0,1]$ into K equally-sized intervals of the form $[i/K, (i+1)/K]$, 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 2. *Assume that the prize is private. Suppose that the collection of all potential initiators 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 a peaceful allocation $\mathbf{x}$. For every group $G \in \mathcal{C}$, we have

$$\int_{j \in G} x(j) \geq v[kp(m) + (1-k)p(m)^2] > vm. \tag{9}$$

Pick a collection of balancing weights $\{\lambda(G)\}_{G \in \mathcal{C}}$. Multiplying each side of (9) by $\lambda(G)$, and adding over all groups in $\mathcal{C}$, we see that

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

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 groups of size $m < m^*$. Then there is no peaceful 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 that a defending group is generally weaker (in a per-capita sense) than its aggressor. 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.

3.3. **Private Prizes and Small Initiators.** With private goods, the intensity of conflict precipitated by small groups is high, because the per-capita payoff (if they do win) is large. To be sure, this does not overturn the fact they have a lower probability of winning than big groups do: in the language of the model, $p(m)$ continues to be increasing in m. But the important point is that the win probability *relative* to group size is high. That fact is reflected in the reverse-logistic shape of the win probability, derived in the proof of Proposition 1. This is why small groups pose a serious threat to peace when the resources at stake are private.

This is not to suggest that small groups will necessarily be the ones to initiate conflict. Proposition 1 asserts that they will be the ones to initiate conflict *if the initial allocation is unbiased*. Of course, if the initial peaceful allocation is chosen to appease the small groups, and there is no stable allocation, 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 the distortion of the peaceful allocation will still have been caused by "over-appeasement" of smaller groups. More generally, the theory is silent on just which peaceful allocation the conflict starts from. But we can say this: the more that overall society is disposed to an equal allocation (and if there is a private prize at stake) the more likely it is that small groups start the conflict. This tendency is *a fortiori* more pronounced if society has a reason to favor larger groups to begin with, as it will in a democratic (or voting) scenario.

Finally, the balancedness condition on initiators, 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.[10] 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 10% make a strict gain from blocking an unbiased allocation. It is possible to check that if there are six such pairwise disjoint group, conflict is inevitable regardless of the baseline allocation: no such allocation can be stable.

The next section paints a different picture when the prize is public.

4. Group Size and Conflict: Public Goods

4.1. **Peaceful Allocations.** Now suppose that the social budget is only used for public goods. Specifically, suppose that there is a one such good for each group (or equivalently, there is 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.

[10] 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.

Once again, the problem of efficient peacetime allocation has an elementary solution. 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 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 unbiased allocation, for instance, each person receives precisely the value Ψm_1.

4.2. **Conflict-Proneness.** As in the case of private prizes, we first study conflict-proneness. If a group initiates and wins a conflict, the use of the grabbed budget is obvious: all of it will go to producing the public good for that group. It follows that the per-capita worth of an initiator (conditional on winning) is just Ψ, and that of the defender (conditional on losing) is zero. If, on the other hand, the defender 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 defender 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 group obtains zero. But this assumption needs more discussion when the prize is public. Suppose that the initiator loses the conflict. It is possible that the defender puts resources into a group-specific public good that some members of the defeated initiator benefit from, simply because they *also* happen to be members of that group (even if they are not members of the defender). 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 suitably taxing them. With this in mind, the per-capita worth of the defender (conditional on winning) is $\overline{\pi} = \Psi\mu$, where μ is the relative size of the largest group that intersects the defender. That is, $\mu = m'/(1-m)$, where m' is the size of that largest group and m, as before, is the size of the initiator. We reiterate that under our approach, this largest group may or may not be fully contained within the defender, but that in either case, the defender appropriates the full surplus (by requiring transfers or reparations, if necessary).

In what follows, consider the (finite) collection of all potential initiators. 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.

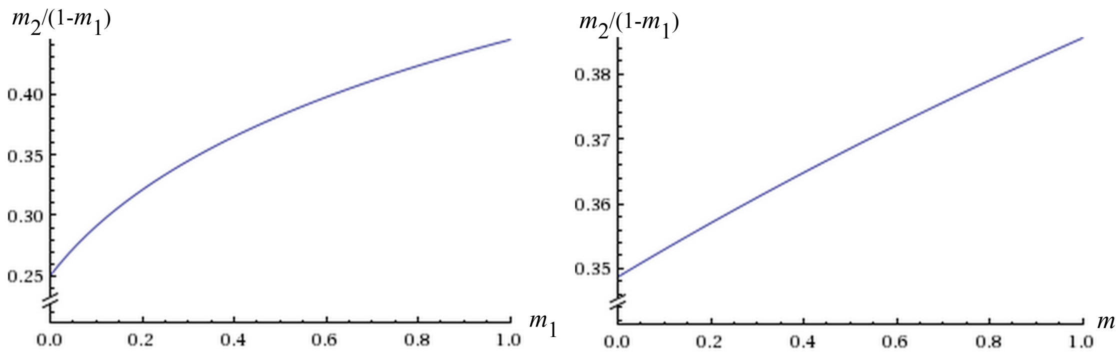
Proposition 3. *Assume that the prize is public. Then society is conflict-prone if and only if*

$$m_1 > \frac{1-\gamma_1 k}{(\gamma_1 - 1)^2}, \tag{10}$$

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

Moreover, provided that (10) applies, the largest group prefers conflict to the unbiased peaceful allocation.

Proof. See Appendix. ■



(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 ($m_2/(1-m_1)$)

The condition (10) 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 (10) blows up, thereby assuring that the inequality cannot be met. With pure public goods, just two groups and full transferability, the unbiased peaceful allocation cannot be blocked. (However, see Section 4.4 on limited transferability.)

Matters are different when there are three groups or more. When an initiator declares conflict against "the rest of society," the defender 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 initiator. 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 (10). 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 group of size 50% or more of the population will want to initiate conflict if the second largest group is around 35% or less of the remaining population.

The situation is more precarious when public goods allocations cannot be fully compensated by financial transfers. For a discussion of this case, see Section 4.4.

4.3. **Actively Conflictual Societies.** We have seen that when the peacetime allocation is unbiased, there are conditions under which society is prone to conflict. With private prizes, we have gone further, arguing that under certain conditions society may be actively conflictual, in the sense that there is no peaceful allocation, unbiased or not, that is immune to attack from some coalition. The same situation can occur with public prizes.

As before, the main question is whether there is some allocation that *simultaneously* avoids conflict from *all* potential initiators. Once again, we need to formulate the idea that the variety of potential threats is large, while avoiding the existence of subgroups that are pivotal for all threats.

Proposition 4. *Suppose that there exists a balanced collection of groups such that for the smallest group s in the collection,*

$$\frac{\eta_s(\eta_s + k)}{(\eta_s + 1)^2} > m_1, \tag{11}$$

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 4. First, as in the case of Proposition 2, a defending group is generally weaker (in a per-capita sense) than its aggressor. In the case of private prizes, that happened because of the Pareto-Olson effect. In the case of public goods, it happens because the defendant 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 2, the defendant 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

$$(n-1)^{1-k} - 2 > (n-1)^k - kn \tag{12}$$

if and only if $n \geq \hat{n}$.[11] The partition is obviously a balanced collection, and condition (17) is the same for every element in it, and can be written as

$$\frac{1}{n} > \frac{1-(n-1)^{1/\alpha}k}{[(n-1)^{1/\alpha} - 1]^2}, \tag{13}$$

because the (common) value of γ is given by $\gamma = (n-1)^{1/\alpha}$ for every group. It is easy to verify that (13) is the same condition as (10), and will therefore hold whenever the number of groups exceeds $\hat{n}$.[12]

Thus both conflict-proneness (in the face of unbiased peaceful allocations) and active conflict (no matter what the peaceful 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 unbiased. The next section shows that the proclivity towards conflict is further

[11] 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 (12) is $k(n-1)^{k-1} - k$, which is strictly declining in n for $n \geq 2$.

[12] Cross-multiply the terms in (13) and divide through by $(n-1)^{1/\alpha}$ to arrive at (12).

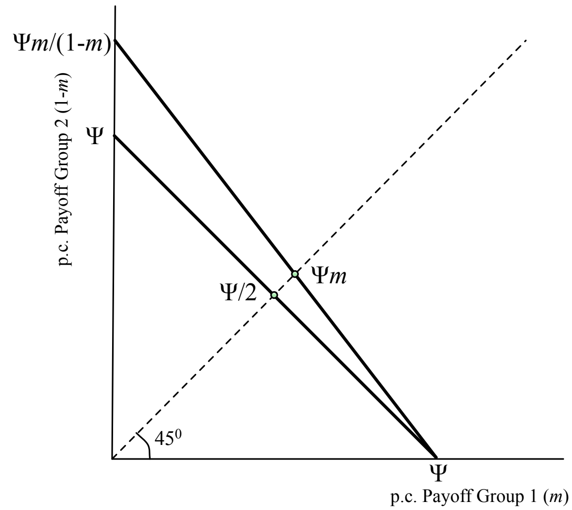


Figure 3. Transferable vs. Nontransferable Payoffs

enhanced if transferability of payoffs is hindered, and as before, large groups will tend to initiate such conflicts.

4.4. **Transferability and Conflict.** 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]

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 unbiased 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 unbiased 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

[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 peaceful payoff frontier is just a finite collection of points. We omit this case for reasons of brevity.

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 (10) 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 unbiased peacetime payoff per person is given by

$$\Psi\left[\sigma m_1 + (1-\sigma)\frac{1}{2}\right], \tag{14}$$

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 5. *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 an initiator of size m and a defendant 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 initiator 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 initiator, 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

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

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

and some straightforward but tedious calculations eventually reveal that

$$m_1 > \left[1 + (n-1)^{1/\alpha} \left\{ \frac{(1+\alpha) - \sqrt{(\alpha-1)^2 + \frac{4\alpha}{n}}}{\sqrt{(\alpha-1)^2 + \frac{4\alpha}{n}} - (\alpha - 1)} \right\} \right]^{-1}. \tag{15}$$

For instance, when there are just two groups and the cost function is quadratic, then the initiating group needs to exceeds 61.8% of the population. When there are three groups and $\alpha = 1.2$, then the initiating group needs to exceed 39.7% of the population. We can use (15) 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.

4.5. **A Remark on Peacetime Allocations and Ethnic Conflict.** This paper is about group size and conflict, and we will now turn to an empirical investigation of that question. But it may be worth mentioning another line of investigation, one that ties in with the question of ethnicity versus class in 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 organised along ethnic lines is a more likely outcome than conflict organised 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 initiate conflict if the prize is private, while initiators are likely to be large 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.

5.1. **Taking the Theory to the Data.** The first empirical question is how to choose the social cleavage (or cleavages) that define 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 suggests channels that could account for the salience of ethnicity (versus class) in conflict. Moreover, the "multiple threats" argument applies even when ethnicity accounts for the *only* set of cleavages in society, as different ethnic groups can pose distinct threats to the State.

The second question concerns the identification of an "initiating group." This is an extremely difficult issue. Even when a conflict is unambiguously described as being started by a particular group, it is hard to fully understand the circumstances that led up to it. In general, that question will have to be dealt with on a case-by-case basis (see, e.g., Mitra and Ray 2014 on Hindu-Muslim violence). Our cross-sectional exercise cannot adequately do justice to this issue. That said, the data we employ, which is a subset of the UCDP/PRIO Armed Conflict Dataset, records *only* conflicts between ethnic groups *and the State*. (For instance, the just-mentioned example of Hindu-Muslim violence will not appear in this database.) That means that the group(s) in question will generally have been involved in some form of demand against a perceived injustice or threat. We certainly do not rule out the possibility of State (or State-sanctioned) violence against that group, but only invoke the assumption that in the vast majority of such cases, that group must have made some political or economic demand or at the very least was actively protecting or promoting its access to economic resources and/or cultural practices.

The third 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).

The fourth question concerns the definition of private and public conflicts. This is another thorny issue: the data are replete with conflicts over private and public payoffs, but in most cases there are mixtures of the two. The Introduction provides some examples of conflicts that can be broadly labeled as being over public or private prizes, but there are always ambiguities. 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. The list goes on.

Fortunately, there is no need for us to take a stand on whether a conflict is public or private; we will allow it have components of either (or both), and our empirical strategies will involve the *interactions* of these components with group size. We adopt (and adapt) a specific approach that we've employed in earlier work; see Esteban, Mayoral and Ray (2012). Briefly, a public prize is at stake if seizing political power brings more benefits, and a private prize is at stake if the homeland of the ethnic group in question is abundant in oil, minerals, and land. While this is a specification that we consistently apply to the entire dataset, it is obviously not comprehensive. For instance, tariff policy can be viewed as a public prize (something that might afford employment protection to a group), but is not studied here as we do not have the data to do it.

In what follows, we introduce the data and the estimation strategy. 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.[15] The dataset contains information for 145 countries and 1475 ethnic groups spanning the years 1960 to 2006.[16]

5.2.1. *Ethnic Groups.* We use the sample of ethnic groups in the "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 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.[17] 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.[18] 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.[19]

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.

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

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

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

[18]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 for distinguishing between two groups in the historic origin of language.

[19]Population figures correspond to the early 60's, see Cederman et al. (2009) for details.

5.2.2. *Conflict.* Data on group-level conflict has been taken from Cederman, Buhaug and Rod (2009), CBR henceforth.[20] 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.[21] We also show that our conclusions are robust to using onset.

5.2.3. *Prizes.* A key prediction of our theory is that the relative size of the initiating group depends on whether the payoff is private or public. In order to test for 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).

Private Prize. To obtain a proxy for the private payoff, 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 ethnic group is rich in natural resources. 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 7.3).

Our baseline measure of group-level oil abundance, OIL, is computed as follows. First, georeferenced information on where oil fields lie and its discovery dates is obtained from PETRODATA (Lujala, Rod and Thieme, 2007). Next, we have combined 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. Results are robust to alternative ways of measuring oil abundance (see Section 7.3).

Public Prize. Our main measure of publicness is very similar to that employed in Esteban, Mayoral, and Ray (2012) who construct a country-level index of "publicness" by asking different questions about the degree of power afforded to those who run the country. The main idea is that the more democratic the country is, the less power the government has, and consequently the lower valuation of the public payoff to conflict. We use five different proxies to construct the index: (i) the degree to which political rights are flouted (POLRIGHTS), (ii) the extent of suppression of civil liberties (CIVLIB), (iii) the lack of executive constraints (EXCONS), (iv) the lack of openness of executive selection (XROPEN) and (v) the lack of competitiveness of executive

[20]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.

[21]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.

selection (XRCOMP).[22] As in Esteban et al. (2012) and Besley and Persson (2011), we don't exploit the high frequency time-variation of the data, since changes in the autocracy variables are likely to be correlated with the incidence of conflict. Thus, our "publicness" measure, PUB, is constructed by first computing time-invariant binary versions of each of the individual indices (in a manner analogous to Esteban et al. 2012), and then averaging the resulting indicators; see Appendix B for details. We have also considered alternative ways of constructing this index; see Section 7.4. In particular, to alleviate concerns of endogeneity due to reverse causality, we have split the sample into two (pre- and post-1980) and have computed the index using pre-1980 data exclusively. Our results are robust to using this index in regressions based on post 1980 information.

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). 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 and year. SOIL-CONST is a measure of the limitations that the group's soil presents to agriculture. 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.** Our baseline specification (corresponding to Column 8 in Table 1) is as follows:

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

for countries $c = 1, \ldots, C$, groups $g = 1, \ldots, G_c$, and dates $t = 1, \ldots, T$. Recall that OIL and PUB 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. Similarly, β_3, the coefficient associated to the interaction of group size and PUB, is expected to be positive, suggesting that the impact of group size on conflict increases as the public prize gets larger.

[22] The variables EXCONS, XROPEN and XRCOMP are three of the five indices employed to compute the "autocracy" index from Polity IV. We don't consider the other two variables — lack of regulation of participation (PARREG) and lack of competitiveness of participation (PARCOMP) — since it has been argued that those components are defined with explicit reference to civil war (Vreeland 2008) and, therefore, should not be included in the type of empirical analysis we are considering here. Thus, our index differs slightly from that employed by Esteban et al. (2012), as they consider all the components of the autocracy variable. Results are robust to using their measure, see Section 7.4.

Dependent Variable: **Conflict Incidence**								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
SIZE	-0.002	0.003	0.008***	0.009***	-0.004***	-0.002	-0.002	0.003
	(0.307)	(0.101)	(0.000)	(0.000)	(0.003)	(0.266)	(0.400)	(0.131)
OIL	0.000**	0.001***	0.001***	0.001***		0.001***	0.001**	0.001***
	(0.040)	(0.009)	(0.002)	(0.006)		(0.008)	(0.024)	(0.009)
SIZE× OIL		-0.001***	-0.002***	-0.002***				-0.001***
		(0.000)	(0.000)	(0.000)				(0.002)
SIZE× PUB					0.010***	0.012***	0.013***	0.010***
					(0.005)	(0.001)	(0.001)	(0.009)
GIP			-0.005***	-0.006***		-0.005***	-0.006***	-0.006***
			(0.003)	(0.003)		(0.003)	(0.003)	(0.002)
GROUPAREA			0.000	0.000*		-0.000	0.000	0.000
			(0.313)	(0.053)		(0.875)	(0.827)	(0.184)
SOILCONST			-0.001*	-0.000		-0.001	-0.000	-0.000
			(0.084)	(0.458)		(0.103)	(0.438)	(0.355)
DISTCAP			0.000***	0.000***		0.000***	0.000***	0.000***
			(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
MOUNT			0.002*	0.002		0.002*	0.002	0.002
			(0.090)	(0.147)		(0.085)	(0.111)	(0.103)
GDP				0.002**			0.001	0.001
				(0.026)			(0.139)	(0.145)
POP				-0.001			0.001	0.001
				(0.795)			(0.509)	(0.559)
LAG	0.895***	0.895***	0.894***	0.892***	0.895***	0.894***	0.893***	0.893***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	0.018***	0.014***	0.009***	-0.020	0.012***	0.010***	-0.042	-0.035
	(0.000)	(0.000)	(0.000)	(0.597)	(0.000)	(0.001)	(0.304)	(0.395)
R^2	0.844	0.844	0.844	0.848	0.844	0.844	0.846	0.846
Obs	64839	64839	64839	53988	64839	64839	57559	57559

Table 1. 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 group level have been computed. p-values are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

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× PUB is changes in ethnic groups within countries, because PUB 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 7.2) we explore this case, as 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.[23]

[23] 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

Thus, for the sake of simplicity we will focus on linear specifications. In Section 7.5, we show that our conclusions remain valid when nonlinear models are employed. Robust standard errors, clustered at the group level, have been computed in all cases.[24]

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. Interestingly, the coefficient of SIZE alone now captures the effect of group size on conflict in the "absence" of a private prize (as measured by oil abundance), and it is positive and significant, suggesting that when there is no oil larger groups are more likely to be involved in conflict. Column 3 (4) adds group (country)-level controls and analogous results are obtained.

Column 5 is similar to column 1 and only includes in the regression SIZE and its interaction with PUB. Observe that PUB cannot be introduced as an 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. The interaction of SIZE and PUB has a positive and significant effect on the probability on conflict. It is also remarkable that the sign of SIZE (that now captures the effect of group size on conflict in the absence of a public prize) is negative and significant, suggesting that smaller groups are more likely to be involved in conflict in the absence of a public prize. Columns 6 and 7 add the group and country-level controls and the interaction maintains both its sign and its significance (although SIZE alone becomes insignificant). Finally, Column 8 (henceforth our baseline specification) includes *both* interactions, which retain their signs and their significance at the 1% level.

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 8 in Table 1. The marginal effect is a function of both OIL and PUB, and as OIL is continuous and PUB is discrete (it takes 5 values), we plot the marginal effect as a function of OIL (in the X axis), for the minimum and

---

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 B.2 for a more detailed discussion.

[24] 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).

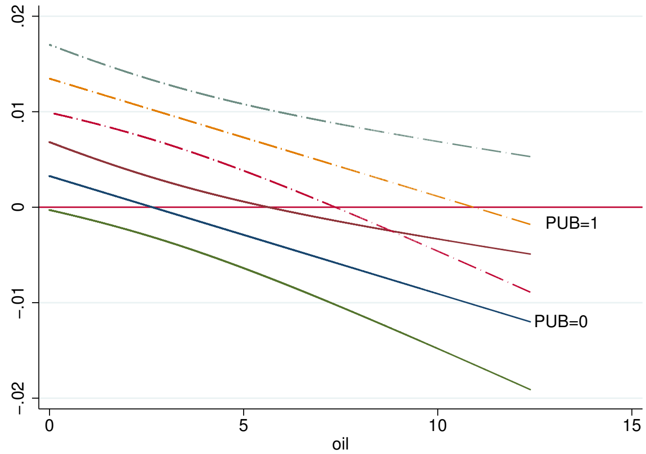


Figure 4. Marginal effect of group size on conflict *incidence* as a function of OIL
This graph depicts the marginal effect of group size on conflict incidence as a function of OIL for two different values of PUB: PUB=0 (bottom line) and PUB= 1 (top line). Confidence bands at the 95% confidence level are also depicted. Estimates from Table 1 (column 8) have been employed to compute the estimates.

maximum values of PUB (PUB=$\{0,1\}$).[25] The solid (dashed) lines correspond to the marginal effect of group size on conflict as a function of OIL together with its confidence bands (at the 95% confidence level) for PUB=0 (PUB=1).

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 PUB 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 PUB 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, it is useful to have an idea not only of the direction of the effect but of its magnitude as well. As the latter also depends on the values of PUB and OIL, we just provide a couple of examples. As the probability that a group is involved in conflict is very small, for ease of interpretation we provide figures relative to the unconditional probability of conflict incidence. For PUB $= 0$ 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 3.2%. Similarly, if OIL $= 0$ and PUB is high ($= 1$, which roughly corresponds to the 80th percentile), an increase of one standard deviation in SIZE increases the probability of conflict by 9%.

[25] 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.

7. VARIATIONS

We now examine the robustness of the baseline results in Table 1. We consider (i) alternative measures of conflict, (ii) group fixed effects, (iii) alternative ways of measuring the private prize, (iv) alternative ways of measuring the public prize, and (v) alternative estimation strategies using nonlinear models.

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 PUB $= 0$ 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 30.2%. Similarly, if OIL $= 0$ and PUB is high, an increase of one standard deviation in SIZE increases the unconditional probability of conflict onset by 75.4%.

7.2. **Group Fixed Effects.** 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. 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$\times$ PUB, given that PUB 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$\times$ PUB). Nevertheless, it is still possible to test one of the two key hypotheses, that pertaining to SIZE$\times$ OIL. Columns 1 and 2 in Table 3 do just that. Note that both columns contain group fixed effects but are still different, because Column 1 excludes lagged conflict while Column 2 includes this variable and is estimated by system GMM (Blundell and Bond 1998). In both cases, the interaction of SIZE and OIL remains negative and significant.

7.3. **Alternative Ways of Constructing the Private Prize.** We've also considered alternative ways of constructing the privateness index. Columns 3 to 6 in Table 3 use alternative definitions of oil abundance. Columns 3 and 4 use OIL(AREA) which is defined as the log of the homeland area covered by oil. The difference between columns 3 and 4 is that the former doesn't include the interaction between SIZE and the public payoff, PUB. The results are very similar to those obtained in the baseline specifications. Columns 5 and 6 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, although when the public prize interaction is introduced as well (column 6), SIZE $\times$ OIL(SHARE) is less precisely estimated (p-value 0.14). Columns 7 and 8 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 7 and 8 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. Interestingly, the interaction of SIZE and

Dependent Variable: **Conflict Onset**								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
SIZE	-0.001	0.003**	0.006***	0.006***	-0.002***	-0.003**	-0.003*	0.001
	(0.333)	(0.025)	(0.000)	(0.000)	(0.003)	(0.043)	(0.081)	(0.633)
OIL	0.001***	0.001***	0.001***	0.001***		0.001***	0.001***	0.001***
	(0.002)	(0.001)	(0.000)	(0.001)		(0.000)	(0.002)	(0.001)
SIZE× OIL		-0.001***	-0.001***	-0.001***				-0.001***
		(0.000)	(0.000)	(0.000)				(0.005)
SIZE× PUB					0.009***	0.011***	0.012***	0.010***
					(0.001)	(0.000)	(0.000)	(0.001)
GIP			-0.003**	-0.003**		-0.003**	-0.003**	-0.003**
			(0.014)	(0.022)		(0.013)	(0.019)	(0.018)
GROUPAREA			-0.000	-0.000		-0.000	-0.000	-0.000
			(0.510)	(0.699)		(0.111)	(0.231)	(0.777)
SOILCONST			-0.001*	-0.000		-0.000*	-0.000	-0.000
			(0.069)	(0.420)		(0.073)	(0.396)	(0.330)
DISTCAP			0.000***	0.000***		0.000***	0.000***	0.000***
			(0.002)	(0.003)		(0.003)	(0.004)	(0.004)
MOUNT			0.002**	0.002**		0.002**	0.002**	0.002**
			(0.023)	(0.049)		(0.019)	(0.039)	(0.036)
GDP				0.001			0.001	0.001
				(0.304)			(0.293)	(0.304)
POP				0.002			0.002	0.002
				(0.270)			(0.233)	(0.259)
PEACEYRS	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	0.099***	0.094***	0.091***	0.009	-0.030***	0.091***	0.004	0.007
	(0.000)	(0.000)	(0.000)	(0.788)	(0.000)	(0.000)	(0.918)	(0.841)
R^2	0.030	0.031	0.031	0.033	0.029	0.031	0.033	0.033
Obs	63187	63187	63187	55611	71136	63187	55611	55611

Table 2. Group Size and Conflict: Onset. *Notes.* 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. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

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.

Table 4 considers mineral availability in the ethnic homeland as a proxy for "privateness." To compute such proxies, we use geo-referenced data on the location of mining activities around the world since 1980.[26] For each year and mine, we have information on whether that mine is active or not, and on the specific minerals produced by it. As in Berman et al (2015), we focus on 13 minerals for which we have world price data,[27] which 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

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

[27]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			0.008***	0.002	0.007***	0.001	0.021***	0.018***
			(0.001)	(0.407)	(0.001)	(0.671)	(0.001)	(0.006)
OIL	0.006	-0.000						
	(0.195)	(0.726)						
OIL(AREA)			0.002***	0.002**				
			(0.008)	(0.011)				
OIL(SHARE)					0.010*	0.010*		
					(0.077)	(0.095)		
AREA(SHARE)							0.023**	0.022**
							(0.022)	(0.032)
SIZE×OIL	-0.011*	-0.001**						
	(0.058)	(0.050)						
SIZE× OIL(SHARE)					-0.023**	-0.011		
					(0.011)	(0.213)		
SIZE× OIL(AREA)			-0.003***	-0.002***				
			(0.001)	(0.010)				
SIZE×(SHARE)							-0.046***	-0.043***
							(0.000)	(0.000)
SIZE× PUB				0.011***		0.012***		0.006
				(0.005)		(0.002)		(0.117)
GIP	-0.031	-0.003*	-0.006***	-0.006***	-0.005***	-0.005***	-0.007***	-0.007***
	(0.172)	(0.052)	(0.003)	(0.002)	(0.004)	(0.004)	(0.000)	(0.000)
GROUPAREA			0.000	0.000	0.000	0.000*		
			(0.442)	(0.554)	(0.176)	(0.054)		
SOILCONST			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.404)	(0.343)	(0.794)	(0.630)	(0.426)	(0.403)
DISTCAP			0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
			(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
			(0.133)	(0.110)	(0.173)	(0.136)	(0.183)	(0.169)
GDP	-0.006	0.004**	0.001	0.001	0.001	0.001	0.001	0.001
	(0.186)	(0.028)	(0.126)	(0.127)	(0.128)	(0.128)	(0.139)	(0.139)
POP	0.016	-0.000	0.001	0.001	0.001	0.001	0.001	0.001
	(0.160)	(0.346)	(0.599)	(0.598)	(0.608)	(0.605)	(0.616)	(0.616)
LAG		0.919***	0.893***	0.893***	0.894***	0.893***	0.893***	0.893***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-0.211	0.000	-0.033	-0.034	-0.035	-0.035	-0.002	-0.028
	(0.322)	(.)	(0.417)	(0.400)	(0.391)	(0.384)	(0.957)	(0.486)
GFE/CFE	GFE	GFE	CFE	CFE	CFE	CFE	CFE	CFE
R^2	0.012		0.846	0.846	0.846	0.846	0.846	0.846
Obs	57559	57559	57559	57559	57559	57559	56756	56756

Table 3. Variations: Group Fixed Effects and Alternative Private Prize Specifications. *Notes.* 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 and 2 use group fixed effects (GFE) and hence only report the interaction SIZE× OIL. Columns 3–6 use alternative oil-based measures of privateness, and Columns 7-8 use land-based measures, as described in the text. Columns 3–8 contain country fixed effects (CFE) and all regressions contain year dummies, and have been estimated by OLS except for Column 2 which is estimated by system GMM. 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 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)

	Dependent Variable: **Conflict Incidence**							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
SIZE	0.010***	0.001	0.011***	0.005	0.010***	0.004	0.010***	0.005
	(0.002)	(0.823)	(0.002)	(0.158)	(0.003)	(0.254)	(0.002)	(0.205)
MINES	0.000	0.000						
	(0.678)	(0.955)						
MINES+OIL			0.000	0.000				
			(0.252)	(0.289)				
SIZE× MINES	-0.002**	-0.001						
	(0.022)	(0.101)						
SIZE× MINES+OIL			-0.002***	-0.001**				
			(0.008)	(0.048)				
MINES(UNWEIGHTED)					0.000	0.000		
					(0.704)	(0.759)		
SIZE× MINES(UNWEIGHTED)					-0.001**	-0.001		
					(0.024)	(0.126)		
MINES+OIL(UNWEIGHTED)							0.000	0.000
							(0.483)	(0.532)
SIZE× MINES+OIL(UNWEIGHTED)							-0.002**	-0.001*
							(0.014)	(0.084)
SIZE× PUB		0.009*		0.009*		0.010**		0.010**
		(0.058)		(0.056)		(0.035)		(0.044)
GIP	-0.005**		-0.005**	-0.005**	-0.005**	-0.005**	-0.005**	-0.005**
	(0.022)		(0.020)	(0.017)	(0.022)	(0.019)	(0.021)	(0.018)
GDP	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	(0.103)	(0.105)	(0.102)	(0.102)	(0.103)	(0.103)	(0.103)	(0.103)
POP	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.579)	(0.605)	(0.588)	(0.583)	(0.574)	(0.572)	(0.574)	(0.573)
GROUPAREA	0.000*	0.000	0.000*	0.000	0.000*	0.000	0.000*	0.000
	(0.059)	(0.127)	(0.062)	(0.105)	(0.059)	(0.102)	(0.062)	(0.108)
SOILCONST	-0.001	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.195)	(0.261)	(0.135)	(0.127)	(0.199)	(0.188)	(0.165)	(0.157)
DISTCAP	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.004)
MOUNT	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001
	(0.429)	(0.295)	(0.413)	(0.387)	(0.430)	(0.404)	(0.428)	(0.403)
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	0.006	0.003	0.004	0.001	0.007	0.003	0.006	0.003
	(0.943)	(0.968)	(0.963)	(0.992)	(0.935)	(0.967)	(0.943)	(0.974)
R^2	0.836	0.836	0.836	0.836	0.836	0.836	0.836	0.836
Obs	35265	35265	35265	35265	35265	35265	35265	35265

Table 4. Variations: Alternative Private Prize Specifications With Minerals. *Notes.* This table regresses conflict incidence on group size and indices of private and public prizes. Different specifications using minerals 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. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

price in 1980 (the year when the 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, particularly when the existence of minerals and oil is jointly considered.

	Dependent Variable: **Conflict Incidence**					
	[1]	[2]	[3]	[4]	[5]	[6]
SIZE	-0.002	0.004	0.000	0.005**	-0.011	-0.001
	(0.407)	(0.129)	(0.919)	(0.010)	(0.206)	(0.882)
OIL	0.001**	0.001***	0.001**	0.001***	0.001*	0.001**
	(0.024)	(0.009)	(0.028)	(0.009)	(0.064)	(0.027)
SIZE× OIL		-0.001***		-0.001***		-0.002***
		(0.001)		(0.000)		(0.002)
SIZE× PUB(PRE SAMPLE)	0.012***	0.009***				
	(0.001)	(0.005)				
SIZE× PUB(EMR)			0.010***	0.009***		
			(0.001)	(0.002)		
RELIGFREEDOM					0.043***	0.043***
					(0.008)	(0.008)
SIZE× RELIGFREEDOM					0.026**	0.022*
					(0.047)	(0.076)
GIP	-0.006***	-0.006***	-0.005***	-0.005***	-0.007*	-0.007*
	(0.002)	(0.002)	(0.003)	(0.003)	(0.052)	(0.050)
GDP	0.001	0.001	0.001	0.001	0.004	0.004
	(0.136)	(0.142)	(0.139)	(0.145)	(0.398)	(0.399)
POP	0.001	0.001	0.001	0.001	-0.009	-0.009
	(0.520)	(0.572)	(0.511)	(0.567)	(0.363)	(0.367)
GROUPAREA	0.000	0.000	0.000	0.000	-0.000	0.000
	(0.966)	(0.215)	(0.907)	(0.154)	(0.207)	(0.933)
SOILCONST	-0.000	-0.000	-0.000	-0.000	-0.001	-0.001
	(0.455)	(0.359)	(0.465)	(0.358)	(0.343)	(0.213)
DISTCAP	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.002)
MOUNT	0.002	0.002	0.002	0.002*	0.000	0.000
	(0.127)	(0.113)	(0.109)	(0.098)	(0.961)	(0.908)
LAG	0.893***	0.893***	0.893***	0.893***	0.832***	0.832***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-0.043	-0.035	-0.042	-0.034	0.063	0.055
	(0.290)	(0.390)	(0.304)	(0.407)	(0.749)	(0.779)
R^2	0.846	0.846	0.846	0.846	0.763	0.763
Obs	57559	57559	57559	57559	22166	22166

Table 5. Variations: Alternative Specifications of the Public Prize. *Notes.* 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. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

7.4. **Alternative Ways of Constructing the Public Prize.** Now we turn to variations on the public prize. In our baseline specification we've used a *time-invariant* index of autocracy, PUB, in order to limit possible high-frequency correlations between conflict and changes in the autocracy variables. Still, reverse causality is a concern as high values in the autocracy index can be the consequence of a history of conflict. To tackle this concern, we have split the sample in two (pre- and post-1980) and have constructed the variable PUB(PRESAMPLE) in the same fashion as PUB but using pre-1980 information exclusively. Columns 1 and 2 in Table 5 use this variable as a proxy for the public payoff in regressions containing only data from the second half of the sample (post-1980). Lagged conflict is maintained as a regressor to forestall the criticism that this variation does not accommodate autocorrelation in conflict. Our conclusions continue to be robust to this variation. Columns 3 and 4 use the measure of publicness employed in EMR (2013); call it PUB(EMR). The difference between PUB and PUB(EMR) is that the former doesn't contain two of the indices included in PUB(EMR) (PARREG and PARCOMP) because it has been argued that they are explicitly defined by political violence or civil war and so might be endogenous (see Vreeland 2008 for details).[28] Using PUB(EMR) instead of PUB yields very similar results.

Finally, Columns 5 and 6 look 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. [29] Column 5 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. Column 6 shows that these results are robust to the introduction of the interaction SIZE $\times$ OIL in the regression.

7.5. **Alternative Estimation Strategies: Nonlinear Models.** Table 6 reproduces Table 1 using a logit specification. The coefficients of the interactions of SIZE and the public and private payoffs maintain the expected signs and are significant at the 1% level in most specifications. 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 B.2 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 theAcholi, 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

[28]The excluded components are: (1) the degree of institutionalization, or regulation, of political competition (PARREG) and (2) the extent of government restriction on political competition (PARCOMP). See Vreeland (2008) for a discussion.

[29]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]	[6]	[7]	[8]
SIZE	-0.177	1.564***	2.760***	2.818***	-3.158***	-2.901**	-2.763**	-1.502
	(0.693)	(0.004)	(0.000)	(0.000)	(0.008)	(0.025)	(0.031)	(0.288)
OIL	0.128***	0.207***	0.257***	0.237***		0.181***	0.170***	0.249***
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
SIZE×OIL		-0.814***	-0.879***	-0.835***				-0.868***
		(0.000)	(0.000)	(0.000)				(0.000)
SIZE× PUB					5.296***	5.664***	5.821***	6.302***
					(0.003)	(0.000)	(0.000)	(0.000)
GIP			-0.681**	-0.681**		-0.657**	-0.663**	-0.609**
			(0.012)	(0.014)		(0.020)	(0.021)	(0.031)
GROUPAREA			0.000	0.000		-0.000	-0.000	-0.000
			(0.864)	(0.459)		(0.372)	(0.368)	(0.775)
SOILCONST			-0.286**	-0.220		-0.176	-0.076	-0.157
			(0.038)	(0.115)		(0.176)	(0.566)	(0.264)
DISTCAP			0.000***	0.001***		0.000***	0.001***	0.001***
			(0.001)	(0.000)		(0.001)	(0.001)	(0.000)
MOUNT			0.554**	0.559**		0.463**	0.474**	0.552**
			(0.016)	(0.020)		(0.041)	(0.044)	(0.022)
GDP				0.270			0.351**	0.353**
				(0.130)			(0.036)	(0.036)
POP				1.589*			1.685*	1.818**
				(0.081)			(0.054)	(0.039)
LAG	7.334***	7.288***	7.216***	7.230***	7.340***	7.240***	7.303***	7.254***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-7.315***	-7.447***	-8.343***	-41.602**	-7.004***	-8.252***	-44.132**	-46.872***
	(0.000)	(0.000)	(0.000)	(0.022)	(0.000)	(0.000)	(0.012)	(0.008)
R^2								
Obs	32776	32776	32776	25572	32776	32776	27344	27344

Table 6. Variations: Logit. *Notes.* This table regresses conflict incidence on group size and indices of private and public prizes. All regressions contain year dummies and country fixed effects, and have been estimated using maximum likelihood in a logit specification. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

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 6 replicates Table 1 using an alternative definition of group size, SIZE-COAL. This variable is defined as follows: for peace years, SIZE-COAL and SIZE coincide. For years where *some* group is in conflict, SIZE-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, SIZE-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.

	Dependent Variable: **Conflict Incidence**							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
SIZE-COAL	0.022***	0.041***	0.059***	0.064***	-0.000	0.012**	0.014**	0.025***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.953)	(0.020)	(0.014)	(0.002)
OIL-COAL	0.002***	0.003***	0.004***	0.004***		0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
SIZE×OIL		-0.005***	-0.005***	-0.005***				-0.003**
		(0.000)	(0.000)	(0.000)				(0.016)
SIZE×PUB					0.067***	0.068***	0.075***	0.070***
					(0.000)	(0.000)	(0.000)	(0.000)
GIP			-0.021***	-0.023***		-0.021***	-0.023***	-0.023***
			(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
GROUPAREA			-0.000***	-0.000***		-0.000***	-0.000***	-0.000**
			(0.002)	(0.002)		(0.000)	(0.002)	(0.018)
SOILCONST			-0.001**	-0.001		-0.001***	-0.001**	-0.001**
			(0.049)	(0.254)		(0.004)	(0.026)	(0.019)
DISTCAP			0.000***	0.000***		0.000***	0.000***	0.000***
			(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
MOUNT			0.003**	0.003		0.004**	0.003**	0.003**
			(0.038)	(0.134)		(0.015)	(0.045)	(0.038)
GDP				0.002*			0.001	0.001
				(0.064)			(0.162)	(0.205)
POP				0.000			0.003	0.003
				(0.885)			(0.131)	(0.193)
LAG	0.886***	0.885***	0.877***	0.874***	0.884***	0.874***	0.870***	0.870***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
c	-0.027***	-0.039***	-0.062***	-0.053	-0.016***	-0.028***	-0.111***	-0.091**
	(0.000)	(0.000)	(0.000)	(0.267)	(0.000)	(0.000)	(0.008)	(0.028)
R^2	0.845	0.845	0.847	0.851	0.846	0.847	0.850	0.850
Obs	64839	64839	64839	53988	64839	64839	57559	57559

Table 7. Group Size and Conflict: Coalitions. *Notes.* 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. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

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 a central 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 initiate conflict when the prize is public, while small groups do so 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.

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

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.

Mitra, A. and D. Ray (2014) "Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India," *Journal of Political Economy* **122**: 719–765.

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.

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 3. 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 $\gamma_1 \equiv [(1-m_1)/m_2]^{1/\alpha}$.[30] Our proof will rely on

Lemma 1. *Assume that the prize is public. Then group i challenges an unbiased 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 group wishes to initiate conflict, so does a larger group, provide both receive their shares of the unbiased allocation.

Proof. Suppose that group 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 initiator and defendant respectively. Using (3),

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

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

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

Overall expected payoffs per-capita to the initiator 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 peacetime allocation to the initiator is unbiased, 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

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

[30] The reason for this asymmetry stems from the fact that the largest group intersecting the defender is of size m_1 for all initiators $i \neq 1$, 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.[31] 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 an unbiased peaceful allocation. Setting $i = 1$ in (17) and manipulating, we obtain (10), as required. ■

Proof of Proposition 4. Let $\mathcal{C}$ be a balanced collection such that (11) 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 $G \in \mathcal{C}$,

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

The remainder of the argument mimics the proof of Proposition 2, 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 $\mathbf{x}$. ■

APPENDIX B

This Appendix contains definitions as well as a table of summary statistics relative to all the variables employed in our empirical analysis. Section B.2 discusses issues relative to the interpretation of interactions in nonlinear models.

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.

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

- 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: this variable 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, Phospate, 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 normalised 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: This variable is 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.): this variable is 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.
- POLRIGHTS: (Lack of) political rights. The data source is Freedom House (2014), which considers a 1–7 scale (1 indicates most free). We transform this variable into a time-invariant dummy in the following way: first, the percentage of years in the sample for which a country received a score smaller than four was calculated. Then, if this percentage was smaller than 40 percent, a country received a value of 1 in all the sample.
- CIVLIB: (Lack of) civil liberties. Data source is Freedom House (2014), which considers a 1–7 scale (1 indicates highest level of liberties). We transform this variable into a time-invariant dummy in the following way: first, the percentage of years in the sample for which a country received a score smaller than four was calculated. Then, if this percentage was smaller than 40 percent, a country received a value of 1 in all the sample.
- EXCONS: (Lack of) executive constraints. It is defined on a 1–7 scale (1 indicates minimum constraints); source is Polity IV (2014). We transform this variable into a time-invariant dummy in the following way: first, the percentage of years in the sample for which a country received a score greater than four is computed. Then, if this percentage is smaller than 0.4, a country received a value of excons equal to 1 in all the sample.

- XROPEN: (Lack of) openness in executive selection. It is defined on a 1–4 scale (1 indicates minimum openness); source is Polity IV (2014). We transform this variable into a time-invariant dummy in the following way: first, the percentage of years in the sample for which a country received a score greater than three is computed. Then, if this percentage is smaller than 0.4, a country received a value of xropen equal to 1 in all the sample.
- XRCOMP: (Lack of) executive constraints. It is defined on a 1–3 scale (1 indicates minimum constraints); source is Polity IV (2014). We transform this variable into a time-invariant dummy in the following way: first, the percentage of years in the sample for which a country received a score greater than two is computed. Then, if this percentage is smaller than 0.4, a country received a value of xrcomp equal to 1 in all the sample.
- PUB: It is defined as the simple average of POLRIGHTS, CIVLIB, EXCONS, XROPEN and XRCOMP.
- PUB(PRESAMPLE): it is identical to PUB but only data pre 1980 has been employed in its construction.
- PUB(EMR): it is constructed in a similar way as PUB but if differs in the indices employed in its construction. PUB(EMR) uses the 5 indices included in the autocracy variable (Polity IV). That is, in addition to XRCOMP, XROPEN and EXCONS, it also includes the degree of institutionalization, or regulation, of political competition (PARREG) and the extent of government restriction on political competition (PARCOMP). These two indices are transformed in time invariant dummies as described above. PUB(EMR) is computed as the simple average of POLRIGHTS, CIVLIB, and the five autocracy indices (EXCONS, XROPEN, XRCOMP, PARREG and PARCOMP).
- 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, 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: It is 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.
- 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.

Summary Statistics

Table B1 presents a statistical summary of the all the variables employed in the empirical analysis.

Table B1. 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
PUB	66096	0.57	0.35	0.00	1.00
PUB(PRE SAMPLE)	66096	0.60	0.33	0.00	1.00
PUB(EMR)	66144	0.51	0.43	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
GDP	56945	7.75	1.16	5.08	11.16
POP	61893	17.08	1.81	11.73	20.98

B.2. **Interactions in non-linear models.** Interpreting the coefficients associated to interactions in linear models is straightforward, as they are simply the cross derivate of the dependent variable with respect to the variables in the interaction. However, this simple logic does not extend to nonlinear models, as shown by Ai and Norton (2003). In non-linear models, the above-mentioned cross-derivative is considerably more involved and important differences arise with respect to the linear case. First, the sign of the interaction does not need to be equal to the sign of the cross derivative; second, its significance cannot be tested with a simple t test on the coefficient of the interaction; and third, its value 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-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, result that mimics the one obtained for the linear case. Figure 6 plots the z-statistics associated to the cross 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

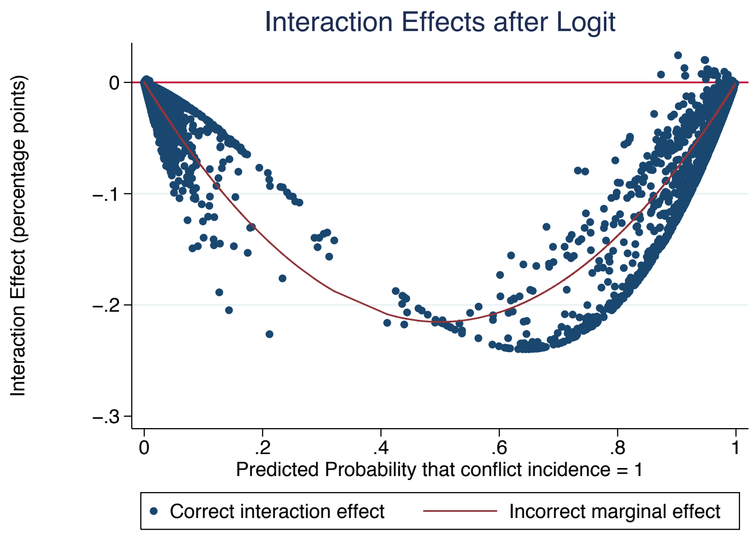


Figure 5. Interpreting interactions in non-linear models
This graph depicts the value of the cross-derivate of conflict incidence with respect to OIL and SIZE for each of the points in the sample. Estimates from Table 6 (column 4) have been employed to compute the estimates.

of the SIZE and PUB: in this case, the cross derivative is positive and significant for most of observations.[32]

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

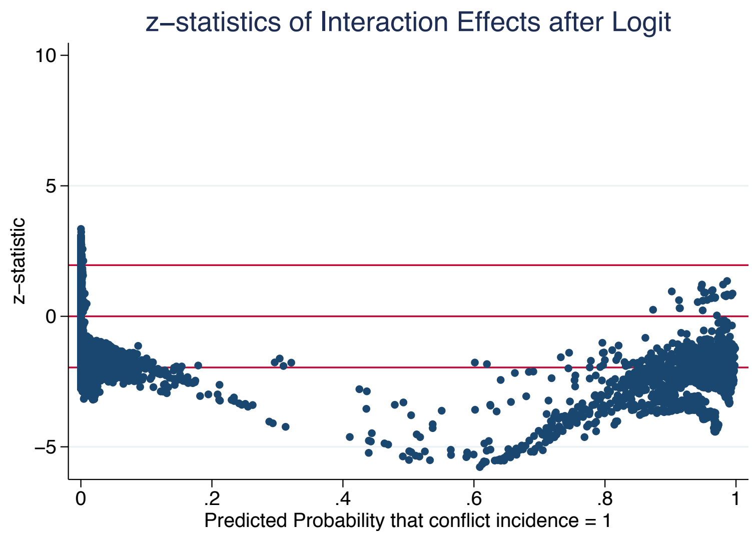


Figure 6. Interpreting interactions in non-linear models: z-statics
This graph plots the values of the z-statistics associated to each of the points of the cross-derivative reported in Figure 5.