Journal of Development Economics 154 (2022) 102759

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

Groups in conflict: Private and public prizes[☆]

Laura Mayoral[a,1], Debraj Ray[b,c,*,1]

[a] Instituto de Análisis Económico (CSIC) and Barcelona School of Economics, Spain
[b] New York University, United States of America
[c] University of Warwick, United Kingdom

ARTICLE INFO

Keywords:
Conflict
Ethnic groups
Oil
Public goods

ABSTRACT

This paper studies costly conflict over private and public goods. Oil is an example of the former, political and civil rights an example of the latter. Our theory predicts that groups in conflict are likely to be small when the prize is private, and large when the prize is public. We examine these implications empirically using a global dataset of conflict at the ethnic group level. Our theoretical predictions find significant confirmation in this setting, and the analysis sheds new light on group size and collective action in the context of violent conflict.

1. Introduction

What is the relationship between the size of an ethnic group and its proclivity to enter into conflict with society? We all know of the "tyranny of the majority" (see, e.g. Tocqueville, 1835), in which a large group more easily imposes its will on society than its smaller counterpart, either through the democratic process or through more coercive means. Yet, there is a contrasting view, a "tyranny of the minority" perhaps, according to which small groups are more involved than large groups in lobbying or conflict, because a given prize goes a longer way *per capita* when groups are small, and also because smaller groups are more cohesive (Pareto, 1927; Olson, 1965; Chamberlin, 1974; McGuire, 1974; Marwell and Oliver, 1993; Oliver and Marwell, 1988; Sandler, 1992; Taylor, 1987; Esteban and Ray, 2001). For instance, Pareto (1927, p.379) writes:

> "[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."

With these perspectives in mind, Fig. 1 displays a binned scatter plot of country-specific ethnic group sizes, and the proclivity of each group to engage in armed conflict in any year, as measured by the share of years between 1960–2006 when they have been so engaged.[2] The estimated relationship, while mildly positive, is statistically insignificant (with a sizeable p-value of 0.555). Empirically, the two "tyrannies" appear to cancel each other. This may be an unfamiliar finding, in part because null findings excite little attention. Nevertheless, it represents the starting point of our paper.

In contrast, our contention is that group size *does* matter, both substantially and significantly, once we properly tease apart the separate contexts in which the two tyrannies might acquire salience. Following Esteban and Ray (2011) and Esteban et al. (2012a,b), we link conflicts to the nature of the disputed surplus. In the theory we develop, groups can challenge some status quo allocation of public or private benefits. A challenge is followed by conflict, with the spoils accruing either to the group – in the case of victory – or to the rest of society, in case the group is defeated. When the prize is private, its per-capita value is affected by group size. When the prize is public at the level of the group, its per-capita value is undiluted by group size. We argue that when an excludable or *private* prize is at stake, a smaller group has a greater incentive to engage in conflict. Conversely, when the dispute is over group-specific *public* prizes, larger groups are associated with conflict.

These observations lead to the two interaction terms – SIZE × PRIVATE PRIZE and SIZE × PUBLIC PRIZE – that we carry through the empirical analysis.

☆ We thank Oeindrila Dube, William Easterly, Joan Esteban and Sahar Parsa for helpful comments. Ray acknowledges support under National Science Foundation grants SES-1629370 and SES-1851758. Mayoral acknowledges financial support from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R+D (Barcelona School of Economics CEX2019-000915-S), La Caixa Foundation Grants on Socioeconomic Wellbeing CG-018-03, and the Ministry of Science, Innovation and Universities Grant number PGC2018-096133-B-100.

* Corresponding author at: New York University, United States of America.
E-mail address: debraj.ray@nyu.edu (D. Ray).

[1] Laura Mayoral and Debraj Ray take equal credit for all the material on the paper.

[2] Country fixed effects are included, along with time-varying controls such as GDP per capita, population and other group-level controls (see Section B.1 of the Online Appendix for a full list).

https://doi.org/10.1016/j.jdeveco.2021.102759
Received 5 February 2021; Received in revised form 4 October 2021; Accepted 9 October 2021
Available online 1 November 2021
0304-3878/

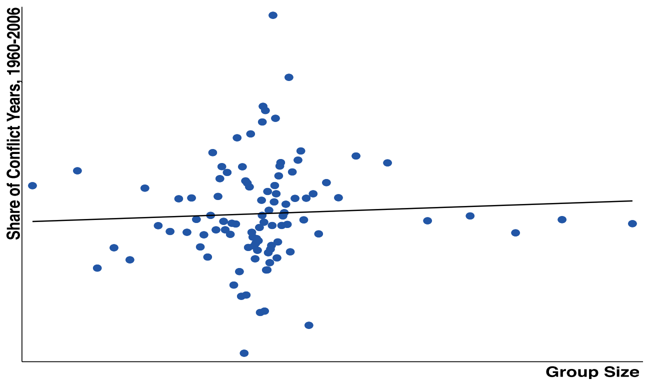


Fig. 1. Conflict and Group Size. This binned scatter plot relates the fraction of years (over 1960–2006) in which a country-specific ethnic group is involved in conflict, to the size of that group. Each point represents 1% of the sample. Several standard controls such as GDP per capita and country fixed effects have been partialed out. There is no significant relationship between group size and conflict (p-value 0.555). Details for this figure are in Section B.1 of the Online Appendix..

In empirical pursuit of our theory, we proxy private prize by the resource endowment on the group's homeland[3] — oil in our baseline analysis, with robustness checks conducted using other resources. We show that the interaction of group size and homeland resources is negatively related to conflict. This interaction effect is substantial enough that the group-size/conflict relationship *itself* turns significantly negative when resources are abundant. In line with the Pareto-Olson thesis, smaller groups are more willing to resist appropriable and excludable endowments.

This relationship is overturned in the presence of a public prize. We proxy the latter by a country-level index capturing the lack of political and civil rights, the idea being that a greater lack of rights signals a larger prize awaiting both a group that is fighting for democratic freedoms, or a group that is itself seeking to seize the reins of autocratic power. (We also conduct several robustness exercises.) We show that group size and conflict are more *positively* related, the larger this public prize. More precisely, the interaction of group size and lack of rights bears a strongly positive relation to conflict.

Fig. 2 illustrates these relationships by displaying binned scatter plots between the share of years in conflict experienced by a group, on SIZE × OIL and SIZE × LACK OF RIGHTS, after employing controls as in Fig. 1. The task of our paper is to carefully develop these connections from both a theoretical and an empirical perspective. In doing so, we gain insight into the channels that can fuel social conflict.

To our knowledge, the interactions between group size and different prizes in precipitating conflict have not been studied. Certainly, group size is implicit in the work of Esteban and Ray (2011), who study the impact of private and public prizes on conflict propensities. But their emphasis is on aggregating these incentives to derive *country*-level measures for empirical implementation. Their theoretical arguments predict that public-prize conflict should be related to polarization (Esteban and Ray, 1994), while private-prize conflict should be tied to fractionalization (see the *Atlas Narodov Mira*; (Bruk and Apenchenko, 1964). Esteban et al. (2012a) explore these predictions empirically. The underlying theoretical underpinnings of these predictions is actually group size, but the effect of group size per se is not explored in these papers.[4] That is the gap we fill here.

Certainly, natural resources – particularly oil – is known to be correlated with conflict; see, e.g., Le Billon (2001), Bannon and Collier (2003), Fearon (2005), Lujala (2010) and Dube and Vargas (2013). Morelli and Rohner (2015) additionally show that the concentration of those natural resources in ethnic homelands is related to conflict. But our focus is on the *interaction* between group size and homeland resources, and not on resources per se. Furthermore, private prizes tell just half the story.

The world is replete with examples of ethnic conflict over both private and group-public prizes. Struggles by ethnic groups seeking to control resources are common (Chechnya, Kashmir, Tamils in Sri Lanka, the Casamance in Senegal, and many other examples). 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). At the same time, the conflict could be a struggle for political control (as in Burundi, Bosnia, Liberia, or Zimbabwe). More generally, a "public prize" could represent (benevolent) democratic freedoms, or the (less benevolent) ability to impose a political or religious ideology, or even the arrogation of private gains that are relatively undiluted by the number of *group* recipients; e.g., public-sector jobs prioritized for a favored ethnicity.[5]

Section 2 introduces the theory. Section 3.1 studies group size and conflict when the prize is private. Section 3.2 does the same for public prizes. Our main empirical results are presented in Section 4. Section 5 contains variations and robustness checks. Section 6 concludes. The Appendix contains detailed definitions of all variables and summary statistics. An Online Appendix contains additional empirical results.

2. Theory

Denote by v the "appropriable resources" of society. These could represent private goods; e.g., oil in a particular location, or public payoffs such as political or cultural power. Suppose that society seeks to allocate v over the entire community. But there is some group, demarcated by ethnicity, geography, religion or occupation, which seeks to retain v for itself. For instance, v might be the value of oil reserves located within the homeland of an ethnic group. The government wants to distribute those revenues over the entire country, while the group might feel that this is "their oil".

The group can accede to a peaceful allocation, or challenge it. Under a challenge, we suppose that society is partitioned into two subsets, the Rebel (our group) and the Defendant (the rest of society, possibly represented by the State). We sidestep the contextual question of whether the Rebel truly initiated the conflict or sought to protect itself against a perceived incursion. For instance, conflict against settlement on Rebel territory may be viewed as self-defense against such aggression. If the Rebel seeks to overthrow the government, then it may be viewed as the aggressor.

Conflict involves the expending of group resources. The utility cost to an individual from a contribution of r is given by $c(r) = (1/\alpha)r^{\alpha}$ for some $\alpha > 1$. Assume that a leader on each side extracts "effort" from everyone to maximize the per-capita payoff of her group.[6] Because the cost of effort is strictly convex, the leader will ask for equal effort from each individual. The winning group obtains full control over the prize.

To map group efforts into win probabilities, we use contest success functions (Tullock, 1980; Skaperdas, 1996). The win probability for the Rebel is given by

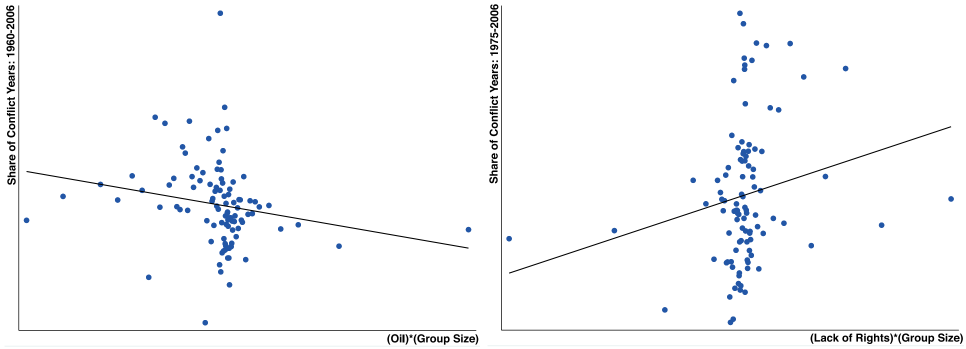
$$p = \frac{mr}{R}, \tag{1}$$

[3] "Homeland" refers to the traditional homeland of the ethnic group, intersected with country.

[4] Esteban and Ray (2001) address the different question of how win probabilities vary with group size in a context of free-riding and both public and private prizes. In the theory here, we get the same results for win probabilities, but we focus on the overall proclivity for conflict.

[5] In this last case, the "prize" has some features of a private good, but as far as the *group* is concerned, it has the dominant features of a club or group-specific public good.

[6] It is easy to write down a variant in which individuals unilaterally provide effort, provided that they at least partially internalize the payoffs of their fellow group members (see Esteban and Ray, 2011).



(a) Size-Oil Interaction **(b)** Size-LackRights Interaction

Fig. 2. Conflict and Group Size: Interactions with Public and Private Prizes. This binned scatter plot relates the fraction of years in which a country-specific ethnic group is involved in conflict, to group size interacted with homeland oil reserves (Panel (a)) or with the lack of political and civil rights (Panel (b)). Each point represents 1% of the sample (around 14 groups). Both graphs partial out the same controls as in Fig. 1. Panel (a) relates the fraction of years (over 1960–2006) in which a group has been involved in civil conflict to group size interacted with homeland oil reserves. The regression estimate is significant and negative (p-value 0.016). Panel (b) relates the fractions of years in conflict (over 1975–2006) to group size interacted with the lack of political and civil rights in that group's country. The regression estimate is significant and positive (p-value 0.014). See Section B.1 in the Online Appendix for additional details on the construction of these graphs.

where r is individual contribution, m is group population share, and $R = mr + \overline{mr}$ is the sum of normalized contributions made by both groups (a bar over a symbol means that the variable pertains to the Defendant). So, if π stands for the per-capita win payoff, and loss payoff is normalized to zero, the Rebel maximizes

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

A similar problem confronts the Defendant, with payoffs $\overline{\pi}$ (or 0). A Nash equilibrium of this "conflict game" is fully described by the first-order conditions:

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

for Rebel and Defendant respectively. Rewrite (2) to observe that

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

for the Rebel, so that its expected payoff from conflict is given by

$$\pi p - c(r) = \pi p - (1/\alpha)\pi p\overline{p} = \pi[kp + (1-k)p^2], \tag{3}$$

where $k \equiv (\alpha - 1)/\alpha$, which lies in $(0, 1)$. This is not in closed form because p, an endogenous outcome, enters this equation. However, (2) implies that:

$$\frac{r}{\overline{r}} = \left(\frac{\pi}{\overline{\pi}}\right)^{1/\alpha} \equiv \gamma, \tag{4}$$

which can be used along with the success function in (1) to conclude that

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

where γ is defined in (4). Together, (3), (4) and (5) describes the Rebel's payoff in conflict equilibrium. A parallel expression holds for the Defendant.

Let $X(v)$ be the set of peaceful allocations. This may range from all possible allocations of v across the population to a very limited set constrained, say, by market outcome or by considerations such as horizontal equity. For instance, a particularly salient peaceful allocation is the *equal* allocation, in which every citizen receives an identical payoff. Say that an allocation $\mathbf{x} \in X$ is *blocked* if the expected conflict payoff to the Rebel exceeds its average peace payoff:

$$\pi[kp + (1-k)p^2] > \frac{1}{m}\int_{i\in\text{Rebel}} x(i). \tag{6}$$

To proceed further, we link the surplus v to the victory payoffs π and $\overline{\pi}$ for each group.

3. Group size and conflict

3.1. Private goods

Suppose the prize is private; say, oil located on the Rebel's homeland. Then $X \subseteq \{\mathbf{x} | \int x(i)di = v\}$. Assume that the winning group seizes the resources v entirely and excludes losers from the spoils. Then, with a Rebel of size m, $\pi = v/m$ and $\overline{\pi} = v/(1-m)$. Using (4), we see that $\gamma = \left(\frac{1-m}{m}\right)^{1/\alpha}$, so that by (5),

$$p = \frac{m^k}{m^k + (1-m)^k}, \tag{7}$$

where $k = (\alpha-1)/\alpha$. Smaller Rebels are disadvantaged in the sense that they have a lower win probability; p is increasing, with $p(1/2) = 1/2$. And yet:

Proposition 1. *If the prize is private, there is $m^* \in (0, \frac{1}{2})$ such that a Rebel with $m < m^*$ will block the equal allocation. Society is conflict-prone with smaller Rebels.*

Proof. The equal allocation gives v to every player. Using (3), conflict payoff is given by $\pi[kp + (1-k)p^2] = v[kp + (1-k)p^2]/m$. So a Rebel of size m will block if

$$kp(m) + (1-k)p(m)^2 > m, \tag{8}$$

where $p(m)$ is given by (7). This function starts above the 45^0 line, crosses it at $n = 1/2$ and then dips below.[7] Fig. 3, which plots p, p^2 and their combination on the left-hand side of (8). This starts out higher than the right-hand side of (8) for small m, but ends up lower. Note that

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

for *any* $m \geq 1/2$.[8] This observation, along with Fig. 3, shows that there is a unique intersection (from above to below) in the interior of $(0, 1/2)$.[9] ■

[7] Note that $\frac{m^k}{m^k+(1-m)^k} \gtrless m$ if and only if $m \lessgtr 1/2$ (simply cross-multiply and verify this).

[8] Suppose this is false for some $1 > m \geq 1/2$. We already know that $m \geq 1/2$ implies $m \geq p(m)$, but then $\text{km}+(1-k)m^2 \geq kp(m) + (1-k)p(m)^2 \geq m$, 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.

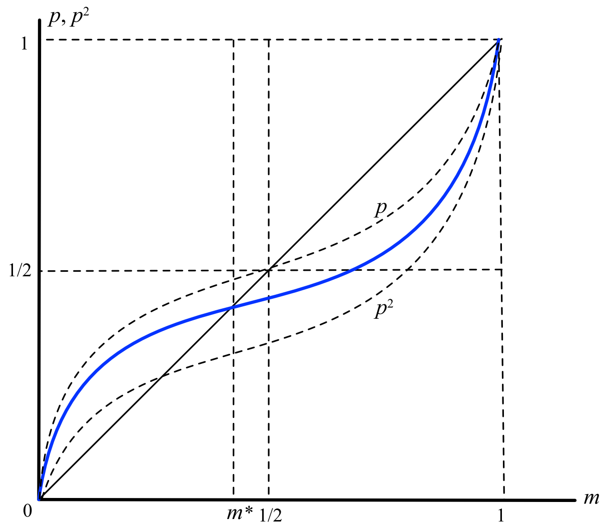


Fig. 3. Threshold for Conflict with Private Prize and Equal Allocation.

Observe that while small Rebels fight more intensely, they do have a lower probability of winning than big groups. Small groups engage in conflict not because they have a high chance of winning. Rather, they do so because they have a high chance of winning *relative* to their share from the equal allocation. That fact is reflected in the "reverse-logistic" shape of the win probability function.

We reiterate that we do not interpret this situation as one in which a Rebel initiates conflict in some unprovoked fashion. State efforts to control resources can be viewed by some group as an unwarranted infringement of its rights, so the salient interpretation could be one of resistance rather than initiation. We also cannot take Proposition 1 as literally applying to *all* group size: surely, some minimum threshold size is needed to pose a threat. We take our model to apply only above such thresholds.

3.2. Public goods

Now suppose that v is a proxy for appropriable *public* payoffs, such as political power, human rights, or funding for secular/religious infrastructure. It could even represent private gains from public policies, such as protectionism or job reservations for an ethnic group.

To fix ideas, suppose that v is a public goods budget allocated to a religious or a secular activity; say, media airtime on national television across religious or secular events. Think of the Rebel as a religious group that benefits from the religious but not the secular activity, and suppose that the rest of society is secular, and does not benefit from the religious activity.[10] Now X consists of the allocations $\mathbf{x} = (x, \bar{x})$ to the Rebel and the rest of society, with $x + \bar{x} = v$, but this time, the per-capita payoffs are also x and $\bar{x}$ to the two groups, and *are not diluted by population*. The equal allocation generates a payoff of $v/2$ per group member, independent of group size. Note the difference from a private prize, where the equal allocation is achieved by dividing the budget in proportion to group population.

In the event of conflict, the Rebel implements its bliss point $(v, 0)$ if it wins, so $\pi = v$. If the Defendant wins, it implements the secular outcome $(0, v)$, so $\overline{\pi} = v$.

Proposition 2. *If the prize is public, there is $\hat{m} \in (0, 1)$ such that a Rebel with $m > \hat{m}$ will block the equal allocation. Society is conflict-prone in the presence of larger Rebels.*

[10] Equivalently, with just two groups, we can normalize these cross-group spillovers to zero.

Proof. In a conflict, (4) implies $\gamma = 1$, so that by (5),

$$p(m) = \frac{m\gamma}{m\gamma + (1 - m)} = m, \tag{9}$$

Applying (9) and (6) to the equal allocation, the Rebel engages in conflict if

$$km + (1 - k)m^2 > 1/2. \tag{10}$$

Define $\hat{m}$ by equality in (10). It is easy to see that $\hat{m} \in (1/2, 1)$, and that for $m > \hat{m}$, (10) holds. ■

With two groups and quadratic costs, we have $k = 1/2$, and using this in (10), we see that the Rebel must exceed approximately 62% of the population. And irrespective of the cost function (10) shows that if the Rebel exceeds approximately 71% of the population, it will contest the equal allocation.

3.3. Some remarks

We collect here some observations on the theory just presented. We do not pursue them here, though they might inform extensions of this research.

First, it is easy enough to combine the two extreme cases of pure private and pure public prizes. Indeed, Esteban and Ray (2011) do just that in the related but distinct context of country-wide distributional measures of ethnic "diversity". The corresponding result here would show a shift in the blocking range from small groups to large groups as the degree of publicness in the prize increases. Specifically, suppose that a fraction $(1-\lambda)$ of the prize is private, while the remaining fraction is public. Then there is $\lambda^* \in (0, 1)$ such that if $\lambda < \lambda^*$, then groups of a certain size or *smaller* would attempt to block the equal allocation, while if $\lambda > \lambda^*$, then groups of a certain size or *larger* would attempt to block the equal allocation.

Second, it is possible to extend these arguments to multilateral conflict. Again, the study of multilateral conflict *per se* has been carried out in Esteban and Ray (2011). The additional concern is to characterize the domino-like manner in which a groups will line up to contest the prize, once some initial group does. Those that do not would need to be viewed as an alliance. This extension is specifically relevant for the public good case, as in practice there will be a number of groups, each with their favorite public good, but possibly with payoff spillovers across groups. The task of "Society" is to reconcile these different claims in some coherent way, and the same overall blocking framework would apply.

A related question concerns the Coase theorem in this setting. Our analysis presumes that peacetime allocations are equal. However, because conflict is inefficient, there is always an allocation that appeases the Rebel. Of course, if the conditions of our propositions hold, that "appeasement allocation" would necessarily need to violate horizontal equity, but at least it Pareto-dominates the conflict outcome, provided that appropriate Coaseian transfers are available. However, under a multiplicity of potential Rebels, there may not be *one* outcome that can *simultaneously* withstand all threats. If the variety of potential threats is large relative to the degree of inefficiency, every peacetime allocation, discriminatory or not, may be blocked by some Rebel. On these matters, see Ray (2010).

On a final and more philosophical note, we need to be especially careful about the transferability of payoffs in the public goods case; specifically, what it means. We presumed that the budgets to produce these goods are transferable across groups, so that "transfers" occur not in units of money but in units of "public goods;" e.g., religion versus secularism. One might want to allow for an additional class of transfers in which compensatory side-payments of money are made from one group to another in exchange for an uneven distribution of public goods. However, such a class should be used with caution, as it presumes that there is a price at which one can compensate for the lack of, say, religious representation, or the sharing of political power. The relative price across objects such as these may be very hard to define.

4. Group size and conflict: Empirics

Our theory implies that smaller groups are more conflictual if the prize is private, and likewise larger groups are more conflictual if the prize is public. There are several considerations that arise when attempting to empirically implement the theory. These include, but are not limited to, a suitable definition of "groups", as well as notions of "private" and "public" payoff components.

4.1. Basic data

We begin with the data that run through the entire exercise.

4.1.1. Ethnic groups

To define potential Rebels, we use ethnicity. Esteban and Ray (2008) provide arguments for ethnic markers to be salient in conflict. Ethnic conflicts account for 50%–75% of internal conflicts since 1945 (Fearon and Laitin, 2003; Doyle and Sambanis, 2006), and so represent a natural choice. We use the sample of ethnic groups from the dataset "Geo-Referencing of Ethnic Groups", or GREG (see (Weidmann et al., 2010). GREG is based on the *Atlas Narodov Mira* or ANV (Bruk and Apenchenko, 1964), created with the aim of locating and charting ethnic groups worldwide. It provides information on the homelands of 929 groups and employs a uniform classification of ethnicity that is consistent across state borders.[11] Most homelands are coded as pertaining to one group, but in some instances up to three ethnic groups share territory. Using this information, we create group-country pairs: that is, we assign ethnic groups to countries depending on the land area they occupy in each country.[12] That yields 1475 distinct group-country pairs located in 145 countries, 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 country population.[13]

GREG's settlement patterns are a snapshot from the late 1950s and early 1960s. That has advantages and disadvantages. On the negative side, settlement patterns are outdated for some parts of the world. Also, as ethnic maps were charted during the Cold War, accuracy and resolution vary considerably for different regions in the world. On the positive side, the use of a snapshot alleviates concerns that ethnic group locations are endogenous to the conflicts we aim to explain. The locational detail in ANV/GREG also enables us to merge it with other geo-referenced datasets needed for the computation of some of our key group-level variables. When all is said and done, we have a panel dataset at the ethnic group level with global coverage, with information for 145 countries and 1475 ethnic groups over 1960 to 2006.[14]

4.1.2. Conflict

Group-level conflict data come from Cederman et al. (2009), who use the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002), checking this against sources that identify ethnic civil wars (such as Fearon and Laitin, 2003; Licklider, 1995; Sambanis, 2001). Ethnic conflicts are coded based on whether mobilization was shaped by ethnic affiliation, and in such cases, the various groups involved are recorded. Our baseline measure is group-level conflict *incidence*, a binary variable set to 1 in any year that group is involved in an armed conflict against the state, with more than 25 battle-related deaths in that year.

One might respond that our theory is about onset, not incidence. But as a matter of pragmatic implementation, it is not possible to trust the onset data as unambiguously pertaining to the start of a genuinely new conflict. As defined by PRIO, group-level conflict *onset* is set equal to 1 in a given year if an armed conflict against the state resulting crosses the 25 battle-related deaths threshold *beginning* with that year. An analysis that literally relies on that definition takes the threshold too seriously. Before it is crossed, we might have several manifestations of serious conflict — a breakdown in negotiations, an insurgency, a crackdown. A year of onset is arguably no different from a year of incidence, though we certainly agree that at a conceptual level, 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, and conclude that with lagged conflict in place, the use of incidence is the better baseline strategy, compared to typing our hands to a far sparser and ambiguous dataset.

That said, we do explore onset as a robustness check, as well as the share of incidence years (and the share of onset years) for which a group has been involved in conflict; see Section 5.2.

4.1.3. Controls

Throughout, we employ a number of group-level controls.[15] These controls are from Cederman et al. (2009) or directly computed from GREG. MOUNT is an index that captures the group's share of mountainous terrain. GROUPAREA is homeland area (in thousands of square km). DISTCAP measures the group's distance to the country capital. GIP is 1 if the ethnic group is in power in the country, lagged one year. SOILCONST measures the limitations of homeland soil for agriculture. PARTITIONED is 1 if the group's homeland is located in two or more countries. LAG is lagged conflict incidence. PEACEYRS is the number of years since the last group-level onset.

4.2. Private payoffs and the effect of group size

We first study group size in a private-prize setting. We consider resources that are located in the homeland of each ethnic group. The underlying presumption is that the State seeks to extract those resources and distribute them more evenly across the country, and that the ethnic group in question can either accept the State policy, or reject it. Alternatively, the group actively seeks conflict to retain those resources. Neither the theory nor the empirics is rich enough to identify the precise perpetrator in these situations, and for our purposes the two are equivalent.

4.2.1. Measures for the private prize

In our baseline, oil in the homeland (OIL) is our proxy for "private prize". First, we obtain geo-referenced information on the location of oil fields and associated discovery dates from Petrodata (Lujala, Rod and Thieme, 2007). Next, we combine group and oil locations from GREG and Petrodata to construct maps of oil fields on ethnic homelands. Finally, OIL is computed as the log of the product of homeland area under oil (km^2) and the international oil price. This measure is intended to reflects the overall value of oil reserves, and so presumes that the average depth of oil across oilfields is similar.[16]

[11] The ANV contains information for 1248 groups, but 319 lack a territorial basis.

[12] An "ethnic group" is not clearly defined in the ANV. It is only possible to infer the coding criteria by comparison with other ethnic databases. Fearon (2003) argues that the main criterion used is the historic origin of language. ANV/GREG contains many small-language groups and more groups overall than alternative sources (e.g., the Geo-Ethnic Power Relations dataset).

[13] Population figures are for the early 1960s; see Cederman et al. (2009).

[14] Morelli and Rohner (2015) consult similar sources for ethnic group location and oil fields.

[15] Since our models contain country-year fixed effects, time-varying country level characteristics are controlled for. However, some specifications use different fixed effects (see Tables B.5, B.7 and B.8). In these models, time-varying country-level controls are also added. More specifically, the log of GDP per capita, lagged one year (GDP), the polity index (POLITY), lagged one year, and the log of total population (POP), also lagged one year. GDP and population are taken from the Penn World Tables while POLITY is taken from Polity IV.

[16] We also consider mineral and land endowments as well as an index of group "resources" (see Table 1), computed via factor analysis on these

4.2.2. Specification

We run the following specification and some variants of it:

$$\text{CONFLICT}_{c,g,t} = \beta_1 \text{SIZE}_{c,g} + \beta_2 \text{PRIV}_{c,g,t} + \beta_3 \text{SIZE}_{c,g} \times \text{PRIV}_{c,g,t} + X'_{c,g,t}\alpha + + \theta_{c,t} + \text{CONFLICT}_{c,g,t-1} + \epsilon_{c,g,t}, \quad (11)$$

for countries c, groups g, and dates t. Our main outcome is "conflict incidence". PRIV is our measure of private prize – OIL will be the leading baseline – its interaction with group SIZE is of particular interest. Our theory predicts that β_3, the coefficient on SIZE × PRIV, is negative, implying that smaller groups are more likely to be in conflict as the private prize becomes more abundant.

Unless otherwise stated, we always employ country-year fixed effects ($\theta_{c,t}$). With these controls, identification for the interaction term SIZE × OIL is achieved via variation in groups within countries (so that SIZE varies) or in known reserves (so that OIL varies).[17] Group level controls ($X_{c,g,t}$) and lagged conflict ($\text{CONFLICT}_{c,g,t-1}$) are also considered.

We estimate Eq. (11) by OLS. We use a linear probability model (rather than a non-linear variant such as probit or logit) because our key variables are interactions and interpreting them in nonlinear models is not straightforward.[18] For completeness, we study nonlinear variants in Section 5.5. Robust and clustered standard errors are computed in all cases. Following Abadie et al. (2017), we cluster errors according to the clustering of the assigned treatment. Whenever the "treatment" of interest is at the group (country) level, we cluster errors at the group (country) level as well. Our results are robust to other clustering strategies such as two-way clustering (at the country and overall ethnic group level). See Table B.4 in the Online Appendix.

4.2.3. Baseline results for private prizes

Each column in Table 1 reports on a different linear probability specification for conflict incidence, with lagged conflict, country-year fixed effects as regressors. Column 1 regresses INCIDENCE on just group size (SIZE) and group-level oil abundance (OIL).[19] The abundance of oil in the ethnic homeland is positively associated with conflict incidence involving that ethnicity. As already observed in the Introduction, this is a well-established correlation. The coefficient on SIZE is insignificant. This is the null finding echoed in Fig. 1. Column 2 introduces the interaction of SIZE and OIL. The coefficient of the interaction term is negative and significant, as predicted by the theory. Column 3 adds group-level controls to the regression in Column 2, with no change in the results.

Columns 1–3 impose the restriction that the marginal effect of SIZE on conflict is linear in oil, and shows that OIL attenuates the effect of size on conflict. But the theory makes a sharper prediction: that the marginal effect of SIZE actually turns *negative* as the prize becomes increasingly private. We could see this by extrapolation of the interaction effect for large private prizes, but that might take our particular specification too literally. Instead, we re-do these Columns using a more flexible specification, one in which linearity is not imposed.

In Column 4, we employ four dummies that correspond to the quartiles of the distribution of OIL for the groups that have oil in their homeland (the omitted category corresponding to groups without oil) and the interactions of group size and each of the oil dummies. Column 4 shows that the latter are negative and significant in groups with large quantities of oil. Columns 5 and 6 employ an alternative way of measuring oil wealth, as they consider national oil rents. Column 5 shows that the same results on the interaction between group size and oil goes through for this case. That said, Column 6 shows that, once the interaction between group oil and group size is introduced in the regression, it is significant, while the corresponding interaction between group size and national oil falls silent. This suggests that, although our conclusions are robust to considering national oil rents, group-level oil seems to be a better proxy for the private payoff.

Columns 7–9 consider other proxies of privateness. These proxies are MINES – mineral availability in the homeland – and HOME, the area of the homeland as a fraction of country area. Both resources can be "seized" for redistribution or settlement. The HOME measure is self-explanatory. For MINES, we use geo-referenced data on global mining activities since 1980.[20] For each year and mine, we know whether that mine is active or not, and the minerals produced by it. Following Berman et al. (2015), we focus on 13 minerals for which we have world price data,[21] taken from the World Bank's commodity price database. For each group, year and mineral, we set a dummy to one if the group has at least one active mine of that mineral. To introduce information on mineral prices, we multiply each of the mineral dummies by (the log of) its international price, normalized by (log) price in 1980. MINES is the sum of the resulting quantities for each group and year.

Column 7 adds the interaction of the mines index and group size; this has a negative and significant coefficient, as predicted.[22] In line with our previous findings, Column 8 shows that the interaction of SIZE and HOME is indeed negative and significant, suggesting that small groups are more likely to be involved in conflict as the value of HOME increases. Finally, Column 9 uses an alternative proxy of privateness, PRIVINDEX, computed using factor analysis on the three prize indicators oil, land and mineral abundance. Similar results are obtained.

How large are these effects? We postpone this discussion to Section 4.3.5, once we have included both private and public prizes in the analysis. That Section also revisits the oil-dummy specification in Column 4, showing that the marginal effect of SIZE on conflict is positive and significant in the absence of oil, but that the effect decreases as oil in the homeland becomes more abundant, *and that it eventually becomes negative* for groups with abundant oil reserves. Our claimed effect is not driven by merely extrapolating a linear specification.

This table – as well as the variations to come – contains an additional observation. Controlling for the SIZE × OIL interaction, SIZE *alone* is generally positive and significant, and its absolute magnitude is decidedly larger than without the interaction. The absolute value of the coefficient of SIZE in Column 3 is almost 4 times larger than that in Column 1. This observation suggests that once a significant private component of the prize is "removed" (i.e., controlled for), group size is positively related to conflict.

endowments: oil, minerals and land. In all cases, we tie private prizes firmly to ethnic homelands. Certainly, the State as a whole can attempt to redistribute the revenues from those resources over the country as a whole, or settle relatively abundant lands with other ethnicities.

[17] With country-year fixed effects, international oil prices also play a role because they affect the value of the (demeaned) interaction for each group and year. Thus, in years with high (low) oil prices, (demeaned) values across groups are larger (smaller) in absolute value.

[18] In linear models, the coefficient of the interaction term is interpretable as the cross derivative of the dependent variable with respect to the variables in the interaction. However, this logic does not extend to nonlinear models: the cross derivative in this case is a more complicated object. As shown by Ai and Norton (2003), its value depends on all the covariates of the model and the sign does not necessarily coincide with the sign of the coefficient of the interaction, see Section B.3.2 of the Online Appendix for a more detailed discussion.

[19] For convenience, the coefficients of SIZE and interactions are multiplied by 10 in all tables.

[20] The source is the *Raw Material Data* (IntierraRMG, 2015). Since data on mining activity starts in 1980, our sample in these regressions focuses on the period 1980–2006.

[21] These are bauxite, coal, copper, diamonds, gold, iron, lead, nickel, platinum, phosphates, silver, tin and zinc.

[22] Similar results hold when only mine availability is used to compute the mines proxy.

Table 1
Group Size and Conflict: Private Prize. This table regresses conflict incidence on group size and indices of private prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain country-year fixed effects. Columns 1–6 employ indices of oil availability, Column 7 uses an index of mine availability, Column 8 uses land-based measures, and Column 9 uses an index of privateness, as described in the text. Robust standard errors clustered at the group level have been computed, except in Columns 5 and 6, as the treatment in these regressions is a country-level variable, see Abadie et al. (2017). Dummy variables OIL_{i-j} equal 1 if the value of oil reserves is between the ith and the jth percentile of the distribution of OIL, conditional on having oil in the homeland. Regressions are estimated by OLS. The time period considered is 1960–2006 except in Column 7, as mine information is only available from 1980 onwards. **p-values** are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Dependent Variable: **Conflict Incidence**								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
SIZE	−0.016	0.028	0.060***	0.071***	0.044**	0.046**	0.054*	0.132***	0.063***
	(0.268)	(0.144)	(0.001)	(0.000)	(0.024)	(0.019)	(0.060)	(0.009)	(0.005)
OIL	0.445**	0.659***	0.806***		0.713**	0.845**	0.564**	0.404*	
	(0.031)	(0.007)	(0.002)		(0.037)	(0.024)	(0.045)	(0.062)	
SIZE×OIL		−12.625***	−14.099***			−46.683**			
		(0.001)	(0.000)			(0.047)			
SIZE × OIL_{0-25}				0.039					
				(0.628)					
SIZE × OIL_{25-50}				−0.040					
				(0.387)					
SIZE × OIL_{50-75}				−0.144***					
				(0.001)					
SIZE × $\text{OIL}_{>75}$				−0.115***					
				(0.000)					
SIZE × OIL COUNTRY					−8.815**	34.495			
					(0.048)	(0.113)			
SIZE × PRIV. IND.									−0.052***
									(0.000)
SIZE × MINES							−0.015**		
							(0.018)		
SIZE × HOME								−0.397***	
								(0.000)	
OIL_{0-25}				−0.001					
				(0.563)					
OIL_{25-50}				0.001					
				(0.697)					
OIL_{50-75}				0.005***					
				(0.002)					
$\text{OIL}_{>75}$				0.006***					
				(0.007)					
MINES							0.000		
							(0.964)		
HOME								0.023**	
								(0.021)	
PRIVATE INDEX									0.002***
									(0.003)
GIP			−0.003**	−0.003**	−0.003	−0.004*	−0.002	−0.005***	−0.004**
			(0.023)	(0.027)	(0.124)	(0.056)	(0.263)	(0.002)	(0.020)
GROUPAREA			0.000	0.000	0.000	0.000	0.000		0.000
			(0.472)	(0.779)	(0.620)	(0.708)	(0.323)		(0.444)
SOILCONST			−0.001*	−0.001*	−0.000	−0.000	−0.001**	−0.001**	−0.001*
			(0.085)	(0.059)	(0.338)	(0.280)	(0.017)	(0.035)	(0.062)
DISTCAP			0.001***	0.001***	0.001	0.001	0.001***	0.001***	0.001***
			(0.000)	(0.000)	(0.153)	(0.122)	(0.003)	(0.000)	(0.000)
MOUNT			0.002*	0.002*	0.002**	0.002**	0.002	0.002*	0.002
			(0.072)	(0.074)	(0.014)	(0.017)	(0.175)	(0.081)	(0.131)
PARTITIONED			−0.000	−0.000	−0.001	−0.001	−0.000	−0.001	−0.001
			(0.570)	(0.586)	(0.495)	(0.416)	(0.928)	(0.494)	(0.530)
LAG	0.904***	0.904***	0.904***	0.903***	0.904***	0.904***	0.892***	0.904***	0.905***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.809	0.809	0.809	0.809	0.810	0.810	0.802	0.811	0.811
Obs	64414	64414	64414	64414	63364	63364	37495	62336	62336

4.3. Public payoffs and the effect of group size

We turn to public prizes. Our baseline has in mind the seizure of political power. That conflict could be a struggle for democracy, in which a group spearheads a movement for broad-based freedoms and rights. Or it could be a struggle for autocratic control, in which a group either seeks the reins of power to favor its own ethnicity, or to impose its religion or ideology on the rest of society. Both are public prizes as far as the *group* is concerned, though in the first case a positive externality is imposed on others, while in the second case the externality is typically negative. Under both conflicts, though, our theory predicts that a larger group is likely to be more active. The spillovers are unimportant for this result: see the very end of Section 3.2 for a discussion of this point.

4.3.1. Measures of the public prize

Our contention is that absence of rights is a good proxy for either notion of the public prize. Under the democratic struggle, the group is at the vanguard of a rebellion that seeks to *restore* political and civil rights, and there is obviously more to restore if there is less of it in the first place. In an autocratic conflict, the absence of political and civil rights is suggestive of the power to be had if an ethnic group can take control of the State. It would be easier to formulate and implement policies that disproportionately benefit members of the group in power.[23] Therefore, in our baseline specification our variable is LACK RIGHTS, a composite index from Freedom House that measures the absence of political and civil rights. It is designed to capture the real-world rights and freedoms enjoyed by individuals. In formulating the index, both the legal guarantees of rights as well as actual practices are taken into account.[24] LACK RIGHTS is the average of two individual indices, the lack of political rights and the lack of civil rights; see Freedom House for additional details.[25] For robustness we will also consider these indices individually (see Columns 3 and 4 of Table 2). We deliberately take LACK RIGHTS off the shelf so as to avoid any implication that the components or weights are chosen to suit our purpose.

We are also aware that there are concerns of endogeneity: for instance, conflict can lead to changes in rights. Therefore, we *only* consider pre-sample values of the index (and in addition we control for past conflict in all our regressions). Specifically, LACK RIGHTS is computed by averaging the values of the Freedom House index from the first year where it is available (which is 1972) to 1975, and is then employed in regressions using post-1975 data only. The resulting measure is "assigned" to all the ethnic groups in the country, so that LACK RIGHTS is a time-invariant country-level index.

We check the robustness of our results by considering other proxies for the publicness index; see Table 2. For instance, we use group exclusion, defined as the average over 1960–1975 of a dummy variable indicating whether a group is excluded from national power (Cederman et al., 2009).

4.3.2. Specification

We add an interaction for public prizes to our specification:

$$\begin{aligned}\text{CONFLICT}_{c,g,t} &= \beta_1 \text{SIZE}_{c,g} + \beta_2 \text{SIZE}_{c,g} \times \text{PRIV}_{c,g,t} + \beta_3 \text{PRIV}_{c,g,t} \\ &+ \beta_4 \text{SIZE}_{c,g} \times \text{PUB}_c + X'_{c,g,t}\alpha \\ &+ \theta_{c,t} + \text{CONFLICT}_{c,g,t-1} + \epsilon_{c,g,t}, \end{aligned} \tag{12}$$

for countries c, groups g, and dates t. The new variable is a public prize index PUB. The theory states that β_4, the coefficient on the interaction PUB × SIZE, is positive. Group-level controls ($X'_{c,g,t}$) as well as country-year fixed effects ($\theta_{c,t}$) are employed, as before. Finally, standard errors have been clustered at the country level, respecting the fact that the "treatment" PUB is generally assigned at the country level.[26]

4.3.3. Baseline results with public prizes

Table 2 summarizes our baseline results for public prizes. It imposes $\beta_2 = 0$ in (12) and so focuses exclusively on the interaction between SIZE and the publicness indices. All specifications contain lagged conflict and country-year fixed effects and are estimated by OLS.

Column 1 contains two regressors: OIL and SIZE while Column 2 adds the interaction between SIZE and LACK RIGHTS, our baseline publicness index. Recall that LACK RIGHTS is a pre-sample time-invariant country-level index, so it is not an independent regressor, and is subsumed in the country fixed effects. But the *interaction* of SIZE and LACK RIGHTS is, of course, well-defined, it has the predicted positive sign and it is highly significant. Column 3 introduces similar controls as in Table 1; the result is robust. Columns 4 and 5 consider separately the two indices employed to compute our baseline measure LACK RIGHTS. These "subindices" measure the lack of civil rights and the lack of political rights, respectively (see Appendix for definitions). Identical results are obtained.

Columns 6 and 7 consider two more proxies of publicness that aim to capture the lack of access to political power. The first proxy (Column 6) is similar to that employed by Esteban et al. (2012a,b) and rests on the idea that there are large gains to seizing power when groups are excluded from it. Specifically, the more "autocratic" a country is, the higher is the value of controlling the State — for both the democratic and autocratic reasons discussed earlier. The measure of lack of access to power that we use is AUTOC, a composite measure of autocracy from Polity IV. As in the case of LACK RIGHTS, we *only* consider pre-sample values of the autocracy index. Specifically, AUTOC averages the values of the autocracy index from 1960 to 1970, which is then employed in regressions using post 1970 data. The resulting measure is "assigned" to all the ethnic groups in the country, so that AUTOC is a time-invariant country-level index.[27] Moreover, we control for past conflict in all our regressions.

The second proxy (Column 7) considers a group-level measure of publicness, based on whether the group is excluded from State power. We construct an index for exclusion, EXCLUDED, in a manner parallel to AUTOC, i.e., by averaging the values of a yearly dummy for exclusion over the period 1945–1970. Identical results are obtained. The interactions of SIZE and the different prize proxies in Columns 6 and 7 keep their expected signs and are significant.

Column 8 employs CHILD MORTALITY as a public prize proxy; see Appendix for the exact definition and sources. We interpret this index as a measure of low provision of services, and take it as a proxy for a large gain under the democratic struggle interpretation. As before, we only use pre-sample values to compute this measure. Column 8 shows that the interaction of group size and child mortality rates is positive and significant; Finally, Column 9 introduces a composite proxy, PUBINDEX, obtained by applying factor analysis on five indicators of (pre-sample) lack of public goods: lack of political rights, lack of civil rights, the level of autocracy, group exclusion from central power and child mortality rates. Our conclusions remain unchanged.

In Section 4.2.3 we observed (as a minor remark) how the coefficient on SIZE turned positive in the presence of the interaction term SIZE × PRIV. In similar vein, and with the interaction SIZE × PUB included instead, the coefficient on SIZE alone turns negative. We only intend this observation as a passing comment, as in no case have we controlled for *all* aspects of public and private prizes, but it is nevertheless suggestive.

4.3.4. Private and public prizes together

Recall that Table 2 imposed $\beta_2 = 0$ in (12) so as to exclusively study the interaction of SIZE and publicness. Table 3 now frees up β_2 and considers simultaneously the interactions of SIZE with indices of privateness *and* publicness. As before, all specifications in Table 3 contain lagged conflict and country-year fixed effects, and are estimated by OLS.

Column 1 shows, with country-year fixed effects but with no other controls, that the interaction SIZE × OIL is negative, the interaction

[23] The resulting individual benefits might be private, but the overall prize – viewed as heightened access to power – is still in the nature of a group-specific public good.

[24] We renormalize the index between 0 and 1, where 0 indicates the lowest rights level.

[25] The political rights index is elaborated taking into account i) the freedom in the electoral process, ii) political pluralism and participation, and (iii) the functioning of the government. The civil rights index evaluates i) the freedom of expression and belief, ii) associational and organizational rights, iii) the rule of law, and (iv) personal autonomy and individual rights. See https://freedomhouse.org/report/methodology-freedom-world-2018 for additional details.

[26] Our results are robust to the consideration of other clustering strategies, such as clustering at the country level or using two-way clustering, with errors clustered at the group (as opposed to country-group) *and* country level. See Table B.4 in the Online Appendix.

[27] See Polity IV for details. The Polity IV manual states: "In mature form, autocracies sharply restrict or suppress competitive political participation ...Our operational indicator of autocracy is derived from codings of the competitiveness of political participation, the regulation of participation, the openness and competitiveness of executive recruitment, and constraints on the chief executive".

Table 2

Group Size and Conflict: Public Prize. Conflict incidence is regressed on group size and indices of public prizes, along with interactions between subsets of these variables as suggested by the theory. Alternative proxies are considered for the public prize. All regressions contain country-year fixed effects, and are estimated by OLS. Robust standard errors clustered at the country level are computed in all Columns, except in Columns 7 and 8, as EXCLUDED is a group-level variable (results are robust to clustering errors at the country level). **p-values** (based on robust standard errors clustered at the country level) are reported in parentheses. For convenience, the coefficients of SIZE ×CHILD MORTALITY are multiplied by 10. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Dependent Variable: **Conflict Incidence**								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
SIZE	−0.017	−0.068**	−0.044	−0.043	−0.042	−0.039	−0.022	−0.047*	0.040
	(0.364)	(0.014)	(0.147)	(0.169)	(0.150)	(0.169)	(0.411)	(0.096)	(0.172)
OIL	0.530	0.587	0.709*	0.706*	0.713*	0.704**	0.553**	0.685*	0.492
	(0.104)	(0.108)	(0.065)	(0.067)	(0.064)	(0.045)	(0.028)	(0.056)	(0.207)
SIZE×LACK RIGHTS		0.091**	0.100**						
		(0.036)	(0.018)						
SIZE×CIVIL RIGHTS				0.105**					
				(0.026)					
SIZE×POL. RIGHTS					0.090**				
					(0.012)				
SIZE×AUTOC						0.116***			
						(0.007)			
SIZE×EXCLUDED							0.111***		
							(0.006)		
SIZE×CHILD MORTALITY								0.006**	
								(0.028)	
SIZE×PUBLIC INDEX									0.035**
									(0.026)
EXCLUDED							0.002		
							(0.506)		
PUBLIC INDEX									0.071
									(0.566)
GIP			−0.003	−0.003	−0.003	−0.003	0.000	−0.003	−0.000
			(0.187)	(0.184)	(0.193)	(0.180)	(0.973)	(0.267)	(0.939)
GROUPAREA			−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
			(0.963)	(0.926)	(0.978)	(0.412)	(0.660)	(0.613)	(0.914)
SOILCONST			−0.000	−0.000	−0.000	−0.001	−0.001*	−0.001	−0.000
			(0.653)	(0.658)	(0.655)	(0.174)	(0.059)	(0.183)	(0.445)
DISTCAP			0.002*	0.002*	0.002*	0.001	0.001***	0.001	0.002*
			(0.079)	(0.080)	(0.078)	(0.144)	(0.001)	(0.137)	(0.078)
MOUNT			0.002	0.002	0.002	0.002*	0.002*	0.002	0.002
			(0.181)	(0.175)	(0.189)	(0.083)	(0.095)	(0.128)	(0.205)
PARTITIONED			−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.001
			(0.675)	(0.676)	(0.677)	(0.659)	(0.818)	(0.734)	(0.533)
LAG	0.901***	0.904***	0.903***	0.903***	0.903***	0.900***	0.901***	0.901***	0.904***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.810	0.817	0.817	0.817	0.817	0.810	0.811	0.809	0.820
Obs	44657	41255	41255	41255	41255	44149	51190	41065	34528

SIZE × LACK RIGHTS is positive and that both are significant. Column 2 adds controls, obtaining identical results. Column 3 shows that our results are robust to considering the flexible specification for oil in Table 1. Column 4 considers oil at the country level (rather than at the ethnic group level and its interaction with group size is negative but insignificant (p-value .12). Column 5 adds the interaction of group-level oil and group size to the specification in Column 4. It is negative and significant, suggesting again that oil based in the ethnic homeland is more relevant than national oil. Column 6 switches the private prize to MINES, or mineral availability in the ethnic homeland. Column 7 switches the private prize again to HOME, the area of the ethnic homeland as a fraction of country area. Column 8 returns to oil as the private prize, while switching the public prize to the AUTOC variable. Columns 9 and 10 consider EXCLUSION and CHILD MORTALITY as alternative public proxies, while retaining OIL for the private prize. It is reassuring that similar results are obtained throughout. These are robust findings.

In keeping with the analysis in Table 3, we note the obvious: while the data are replete with conflicts over private and public payoffs, the two are often intertwined. For instance, even a conflict as seemingly primordial as Rwanda was permeated with economic looting, such as land grabs, under the cover of ethnic violence. The Second Civil War in the Sudan is about different cultural and religious identities, but it is also – to some degree – about oil; so is the Chechnyan War. The Zimbabwean conflict is about identity and political power, but it is also about land, and so on. In the light of these expected complications, it is of interest that the two interaction predictions made by the theory hold up separately and robustly across different variations.

4.3.5. *Coefficient magnitudes*

With both interaction effects in place, we are in a position to provide a sense of the magnitudes of the estimated coefficients. We use the estimates from Column 2 from Table 3, which contain our two baseline indices for public and private prizes along with other controls. The estimated marginal effects of group size, coming as they do from interactions, must depend on the values of LACK RIGHTS and OIL, so we give a couple of examples here and refer the reader to Fig. 4 for more. For LACK RIGHTS = 0 and a high value of oil (at the 95th percentile) an increase of one standard deviation in SIZE decreases the unconditional probability of conflict incidence by 4.03%. Similarly, if OIL = 0 and LACK RIGHTS is high (= 1), an increase of one standard deviation in SIZE increases the probability of conflict by 5.5%.

We have seen that the presence of oil attenuates the effect of size on conflict. But as already mentioned, the theory makes a sharper prediction: the marginal effect of SIZE actually turns *negative* as the prize becomes increasingly private. (The opposite is true when the prize is public.) The two examples in the previous paragraph are provided with this in mind. More generally, Fig. 4 plots the marginal effect of SIZE on INCIDENCE computed using the estimates from Column 2 in Table 3. The marginal effect is a function of both OIL and LACK RIGHTS, and the plot displays this marginal effect as a function of OIL (on the x-axis), for the

Table 3

Group Size and Conflict: Private and Public Prize Specifications. This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain country-year fixed effects, and have been estimated by OLS. **p-values** (based on robust standard errors clustered at the country level) are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Dependent Variable: **Conflict Incidence**									
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
SIZE	0.001	0.025	0.034	−0.008	0.001	0.010	0.116*	0.023	0.036	0.024
	(0.973)	(0.344)	(0.223)	(0.736)	(0.972)	(0.736)	(0.087)	(0.411)	(0.204)	(0.395)
OIL	0.749*	0.845*		0.714*	0.900**	0.550	0.431	0.801**	0.752***	0.847**
	(0.093)	(0.051)		(0.081)	(0.043)	(0.163)	(0.198)	(0.037)	(0.007)	(0.042)
SIZE × OIL	−13.410**	−15.432**			−55.681*			−13.998**	−14.906***	−13.668**
	(0.045)	(0.022)			(0.060)			(0.024)	(0.000)	(0.039)
SIZE× LACK RIGHTS	0.067*	0.078**	0.079*	0.092**	0.078**	0.079*	0.050			
	(0.072)	(0.036)	(0.056)	(0.022)	(0.034)	(0.082)	(0.258)			
SIZE × OIL_{0-25}			−0.008							
			(0.947)							
SIZE × OIL_{25-50}			0.306							
			(0.472)							
SIZE × OIL_{50-75}			−0.164*							
			(0.052)							
SIZE × $\text{OIL}_{>75}$			−0.133***							
			(0.008)							
SIZE× OIL COUNTRY				−7.531	43.541					
				(0.127)	(0.111)					
SIZE × MINES						−0.015*				
						(0.067)				
SIZE × HOME							−0.418**			
							(0.018)			
SIZE × AUTOC								0.100**		
								(0.014)		
SIZE × EXCLUDED									0.100***	
									(0.010)	
SIZE× CHILD MORTALITY										0.004
										(0.142)
OIL_{0-25}			−0.002							
			(0.398)							
OIL_{25-50}			−0.002							
			(0.505)							
OIL_{50-75}			0.006**							
			(0.048)							
$\text{OIL}_{>75}$			0.006**							
			(0.032)							
MINES						0.000				
						(0.887)				
HOME							0.023*			
							(0.075)			
EXCLUDED									0.003	
									(0.178)	
GIP		−0.003	−0.003	−0.003	−0.004*	−0.002	−0.004*	−0.003		−0.003
		(0.191)	(0.187)	(0.229)	(0.098)	(0.310)	(0.057)	(0.189)		(0.269)
GROUPAREA		0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000
		(0.166)	(0.153)	(0.410)	(0.261)	(0.256)		(0.685)	(0.293)	(0.589)
SOILCONST		−0.000	−0.000	−0.000	−0.000	−0.001	−0.001	−0.001*	−0.001**	−0.001
		(0.472)	(0.442)	(0.544)	(0.480)	(0.352)	(0.334)	(0.086)	(0.027)	(0.121)
DISTCAP		0.002*	0.002*	0.002*	0.002*	0.002	0.002*	0.001	0.001***	0.001
		(0.085)	(0.080)	(0.096)	(0.061)	(0.110)	(0.076)	(0.139)	(0.001)	(0.128)
MOUNT		0.002	0.001	0.002	0.001	0.002	0.002	0.002*	0.002*	0.002
		(0.180)	(0.187)	(0.182)	(0.220)	(0.270)	(0.214)	(0.067)	(0.079)	(0.114)
PARTITIONED		−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
		(0.755)	(0.704)	(0.753)	(0.697)	(0.927)	(0.717)	(0.713)	(0.813)	(0.739)
LAG	0.902***	0.902***	0.901***	0.902***	0.901***	0.895***	0.902***	0.899***	0.901***	0.901***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.818	0.818	0.818	0.818	0.818	0.809	0.820	0.811	0.811	0.809
Obs	39969	39969	39969	39969	39969	34639	38689	42757	51258	41065

minimum and maximum values of LACK RIGHTS (i.e., LACK RIGHTS$=\{0,1\}$). The dashed lines represent 90% confidence bands. In line with the theory, the figure shows that the marginal effect of size is negative or positive, depending on the values of the public and private payoffs. For a small value of LACK RIGHTS and moderate or large values of OIL, an increase in group size has a negative and significant effect on conflict incidence. The opposite is true when LACK RIGHTS is high and OIL is small; now 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. Table B.1 in the Online Appendix makes these points in an even simpler way by taking binary cuts for private and public prizes and comparing the four cells that are then generated.

Recall that our baseline specification imposes the restriction that the marginal effect of SIZE on conflict is a linear function of oil. One might therefore argue that the above observations stem from that assumed linearity. However, Column 3 in Table 3 shows that similar results are found when a more flexible specification is employed, one in which linearity is not imposed. Recall that this Column uses four dummies that correspond to the quartiles of the distribution of OIL for the groups that have oil in their homeland (thus, the omitted category corresponds to groups that do not have oil). In this case, the marginal

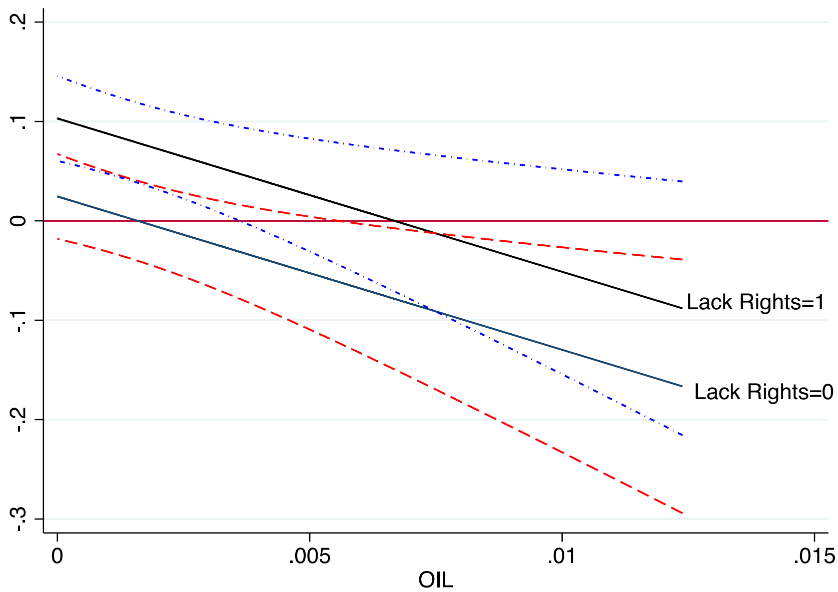


Fig. 4. Marginal Effects of Group Size on Conflict. This graph depicts the marginal effect of group size on conflict incidence as a function of OIL for two different values of LACK RIGHTS: 0 (bottom solid line) and 1 (top solid line). Confidence bands at the 90% level are also depicted. Estimates from Table 3 (Column 2) have been employed to construct the graphs.

effect of SIZE on conflict is given by the sum of the coefficient of SIZE and that of the variables SIZE × OIL_j, where OIL_j is equal to 1 if the group's oil is in quartile j. The marginal effect of SIZE on conflict is positive and significant in the absence of oil, but the effect decreases as the amount of oil in the homeland becomes more abundant and it eventually becomes negative for groups with abundant oil reserves. In particular, using the results in Column 3, we can reject the assertion that the sum of the coefficient of SIZE and that of SIZE× OIL_{50-75} (or SIZE× $\text{OIL}_{\geq 75}$) is greater than or equal to zero.[28] So this effect is not driven by merely extrapolating a linear specification.

The negative relationship between size and conflict under a private prize – morphing into a positive relationship as the prize turns public – is a central finding of the theory and the empirical analysis. Section 5.1 argues that it is hard to think of an alternative explanation that generates the same joint pattern.

5. Variations

The evidence so far shows a robust link between the probability of conflict, group size and the nature of the payoffs. In this section, we consider several variations on the baseline exercises. First, we consider alternative explanations that could rationalize our empirical findings and provide evidence against them (Section 5.1). We then go on to consider alternative measures of conflict (Section 5.2), and the possibility that our results are due to omitted variables more generally, finding little support for this (Section 5.3). We end with some statistical variations: two-way clustering of standard errors instead of country- or group-level clustering (Section 5.4), the use of a nonlinear model; specifically logit (Section 5.5), robustness to dropping different regions of the world (Section 5.6), and group and year fixed effects, as well as country and year fixed effects (Section 5.7). The corresponding tables of results (labeled with the prefix B) are provided in the Online Appendix.

5.1. Alternative explanations

One possible interpretation of our results invokes differences in conflict technology rather than the structure of individual payoffs generated by group sizes and the nature of the prize. According to this view, large groups (with or without oil) would have easier access to the funds needed to engage in conflict against the State. However, small groups would find it particularly useful to have oil in their homeland to purchase weapons, hire mercenaries, etc., which otherwise would be beyond their means. As in the case of our theory, this explanation would generate a heterogeneous impact of group size on conflict: small groups would tend to fight less than large groups, unless they have oil. However, this alternative explanation would fail to generate the *negative relationship* between size and conflict shown in Tables 1 and 3 and in Fig. 4. While the effect of size on conflict would be attenuated by oil, the net effect of group size must always remain positive.

Morelli and Rohner (2015) study the relationship between conflict and the *concentration* of natural resources in ethnic homelands. They show that the larger a group's share of oil, the larger the probability of conflict onset. This is a completely different prediction from ours; it is orthogonal to what we do. That said, and because bigger groups are more likely to have a larger share of national oil, we check that oil share is not a confound in our regressions. Columns 1 and 2 in Table B.2 add to our baseline specifications the share of oil as computed by Morelli and Rohner (i.e., the surface of an ethnic group's territory covered with oil and gas as a percentage of total country surface covered with oil and gas). Oil share is far from being significant in these specifications. Our conclusions survive unchanged, although the coefficient of SIZE × OIL is estimated more noisily in Column 2 (p-value is .101), when the public prize interaction is also in the regression.[29]

Our theoretical results stress the fact that if the initial allocation is equal – or if it is biased *against* small groups – then the latter are more likely to be involved in conflict if the payoff is private. An alternative

[28] p-values are .070 and .006, respectively.

[29] Morelli and Rohner (2015) consider onset rather than incidence. We have also checked that in onset regressions the significance of the share of oil vanishes once one controls for total group oil, but again, this is not our focus.

interpretation, however, would run as follows: large groups are stronger and, as a result, more likely to be in power. Thus, they "automatically" get a large share of the rents from the center, and so are less likely to rebel. Although related to our argument, the underlying mechanism is different: small groups rebel because they are more likely to be excluded from power and so do badly under the initial allocation. This alternative explanation is, however, at odds with other aspects of our empirical results: in that case we should see small groups fighting more, not just on account of oil, but simply because they are treated worse. But there is no evidence of that at all: in *all* our specifications, group size per se is either insignificant or positive whenever significant.

Absent a direct measure of the initial allocation, we do control in all our regressions for whether a group is included or excluded from political power. In addition, we have also considered whether our results are robust when only *excluded* groups are considered.[30] Columns 3 and 4 in Table B.2 restrict the sample to groups excluded from power and shows that our conclusions continue to hold when only those groups are in the sample.

Similarly, the index LACK RIGHTS is typically high in authoritarian regimes, which often tend to have a ruling elite made up of minorities. So it would be possible that in those regimes conflict is initiated by majorities that want to take over power from these minorities. Indeed, the average size of the group(s) in power in autocracies is smaller than in less autocratic regimes. To rule out this possibility, we drop countries whose ruling elites are small (as compared to ruling elites in non-autocratic countries). More specifically, we divide the sample into autocratic and non-autocratic countries (defined as those with autocracy index higher/lower than 5) and we drop from the sample autocratic countries where the size of the ruling elite is smaller than the median of the size of the group(s) in power in non-autocracies. Then we re-run our baseline specification with this reduced sample. Our results remain robust to this variation, see Columns 5 and 6 in Table B.2.

Recall our remark at the end of Section 4.2.3: that SIZE becomes positive and significant after its interaction with OIL is controlled for. Columns 8–10 explore our result with a similar perspective in mind. The theory rests on the idea that large groups are less likely to fight because payoff per head is relatively low. The implication is that if one controls for *per capita* payoffs (rather than for total payoffs, as in the specifications above), large groups should *unambiguously* be more conflict-prone. To explore this prediction, Columns 8–10 control for per capita private payoffs. To facilitate comparison with the results so far, Column 7 controls once again for OIL, our baseline measure of total private payoff, and shows that group size is insignificant. Column 8 is identical to Column 7 but replaces OIL by OIL PC, computed by dividing OIL by group population. In this case, the coefficient of SIZE is significant and more than doubles that in Column 7, suggesting that larger groups are more prone to conflict once per capita payoffs are held constant. Columns 9 and 10 show that a similar result also holds when land per capita (Column 9) and land and oil per capita (Column 10) are introduced in the regression.[31]

Finally, one could, of course, posit something with no particular conceptual foundation: that oil is special for smaller groups, or that small groups have a better conflict technology, or that small groups are particularly fond of secession and fight harder for their freedom. These *ad hoc* alternatives must all contend with the simple observation explored in Columns 8–10 of Table B.2: that controlling for the *per-capita value* of the prize, group size is positively related to conflict. None of these explanations can also explain why the relationship turns positive when per-capita controls for private wealth are imposed.

The one reasonable (and related) argument that does generate a negative relationship between group size and conflict in the presence of private goods is the free-rider argument first described by Olson (1965). Small groups are better capable of cohesion. As already discussed, this argument complements the one based on per-capita payoffs first described by Pareto (1927), that we emphasize in our theory. The problem is that the free-rider theory works well for one side of our observations but not the other. It would have no prediction for group size and *public* prizes, where the observed relationship is positive.

Apart from the evidence against the above arguments, it is to be noted that these arguments pertain to pieces of our main result and not the entire prediction. The negative relationship between size and conflict under a private prize – turning into a positive relationship as the prize becomes public – is our central finding. We could not think of an alternative explanation with the same *joint* pattern.

5.2. Alternative measures of conflict

Table B.3 in the Online Appendix considers alternative measures of conflict: Columns 1–3 use conflict *onset* in a panel set-up. Columns 4–8 drop the time dimension of the data: the dependent variables are the share of years over 1975–2006 in which a group has been involved in conflict against the State (Columns 4–6), and the share of onset years (Column 7–8), as in the Introduction. 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 generally highly significant.

5.3. Assessing the importance of the omitted variable bias

Despite our attempts to control for a large number of potential confounders, we still cannot completely rule out the possibility that unobserved variables are biasing our results. However, it is possible to assess the likelihood that our observed effect is solely due to selection bias. To that effect, we apply a technique recently developed by Oster (2019), which builds on the work by Altonji et al. (2005) and Bellows and Miguel (2009). This method allows to determine the degree of selection on unobservables relative to observables (denoted by δ) that would be necessary to explain away the result. If the set of observed controls is representative of all possible controls, then a large value of δ suggests that it is implausible that omitted variable bias explains away the entire effect. Altonji et al. (2005) and Oster (2019) suggest the use of a cut-off of δ (δ^*) equal to 1. This value means that selection on unobservables would need to be at least as important as that on observables to produce a treatment effect of zero. Thus, if for example a value of $\delta = 2$ is obtained, it would suggest that the unobservables would need to be twice as important as the observables to produce a treatment effect of zero. One reason to favor the cut-off $\delta^* = 1$ is that researchers typically choose the controls they believe ex ante are the most important (Angrist and Pischke, 2010) and thus situations where selection on unobservables is larger than that of the observed controls are deemed unlikely.

The statistic employed to compute δ is designed to evaluate the stability of the variable(s) of interest to the introduction of controls. More specifically, it is a function of the coefficient of the variable of interest estimated in a full model (that contains all controls), the same coefficient obtained in a restricted model with no (or few) controls, the R^2s obtained in these regressions and $R^2_{\max}$, the R^2 from a hypothetical regression of the outcome on treatment and both observed and unobserved controls. If the outcome can be fully explained, then $R^2_{\max} = 1$. However, as acknowledged by Oster (2019), in most empirical settings it seems likely (due, for example, to measurement error) that the outcome cannot be fully explained even if the full control set is included.

In our case, the variables of interest are either SIZE × OIL or SIZE × LACK RIGHTS. The full model corresponds to the specification containing

[30] More specifically, we drop from the sample groups with a value of EXCLUDED (a pre-sample average of an exclusion dummy over the years 1960–1975) larger than 0.5. Results are very robust to other ways of defining exclusion.

[31] We cannot do the same with MINES as we do not have a measure of total production.

both interactions and all controls (Column 2 in Table 3). The restricted model is one with no controls.[32] The values of δ are quite sensitive to the choice of the (unobserved) value of $R^2_{\max}$. We compute the maximum value of $R^2_{\max}$ we can use in order to obtain values of δ larger than 1. This value turns out to be quite large (around 1 and 0.93 for the private and public interaction, respectively). Since the variables employed in our regressions are clearly not perfectly measured, we believe that this is a reasonable value for the maximum R^2 that can be achieved. Therefore, our findings give support to our claim that omitted variable bias is unlikely to explain our results.

5.4. Alternative clustering strategies

As explained in Section 4.2.2, we follow Abadie et al. (2017) and cluster errors at the group or at the country level, depending on whether the corresponding treatment is assigned at the group or at the country level. Our results are firmly robust to other clustering strategies; for instance, to two-way clustering. Table B.4 recomputes our baseline regressions but this time standard errors are adjusted for clustering at the ethic homeland *and* at the country level. Notice that since ethnic homelands are often split by an international border, the latter dimensions are not nested. Our conclusions are robust to considering alternative clustering schemes.

5.5. Nonlinear models

We re-estimate our baseline specifications using logit. Table B.5 in the Online Appendix presents the results. Country-year fixed effects have been replaced by country and year fixed effects, as otherwise computational costs are too high. All equations contain the full list of group-level controls employed in the previous Tables and add three time varying country-level: the log of GDP per capita, the log of total population and Polity. Columns differ on the interactions included in them: Column 1 includes the interaction of SIZE and OIL, Column 2 that of SIZE and LACK RIGHTS, while Column 3 considers both of them. The coefficients of the interactions of SIZE and the public and private payoffs maintain the expected signs and remain significant. In nonlinear specifications, however, one has to be cautious when interpreting the change in two interacted variables, as Ai and Norton (2003) pointed out. The Online Appendix discusses this issue in more detail, and argues that our conclusions still hold.

5.6. Dropping regions of the world

Table B.6 in the Online Appendix drops observations from different regions: former USSR countries (Columns 1 and 2), Asia (Columns 3 and 4), Middle East (Columns 5 and 6), West-South Africa (Columns 7 and 8), East and Central Africa (Columns 9 and 10), and Latin America (Columns 11 and 12). For each region, the first (second) Column consider specifications with and without the interaction of group size and the public prize. Results are generally robust, except when East and Central African observations are dropped. The interaction of SIZE and OIL is still significant in Column 9, but when SIZE × LACK RIGHTS is introduced as well (Column 10), it ceases to be so (the *p*-value of the public interaction is 0.16). To put this result in perspective, however, notice East and Central Africa is by far the most conflictual region in our sample:, with 15% of all observations but 30% of all conflict observations. Unsurprisingly, the coefficients are estimated less precisely when many observations are dropped.

5.7. Other specifications for fixed effects

Table B.7 replaces the country-year fixed effects considered in all regressions by group (and year) fixed effects. The inclusion of group fixed effects contributes to the reduction of potential bias from omitted variables, which is a good thing. The drawback is that all time-invariant controls drop out from the regression, including one of our key variables, SIZE. Nevertheless, it is still possible to test the key hypothesis pertaining to SIZE × OIL. With group fixed effects, the identification for the interaction term SIZE × OIL is achieved via variation in the value of OIL over time, either because of the discovery of new reserves or due to fluctuations in oil prices. Finally, Table B.8 reproduces Table 3 but using separate country and year fixed effects. Tables B.7 and B.8 show that our results continue to hold when alternative sets of fixed effects are considered.

6. Conclusion

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 … 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. We discussed their contributions in detail above.

In this paper we have studied a model of social conflict, in which the conflict may be over a *public* or a *private* good. The main result, that we explore empirically through a variety of specifications, is that conflict is more likely in the presence of a private prize when the group is small, and it is more likely in the presence of a public prize when the group is large. By using a global panel dataset at the ethnic group level we find powerful and robust empirical support for these claims. This is our approach to reconciling Tocqueville–Mill with Pareto-Olson.

Our approach can be extended to other questions of interest. Specifically, as suggested in Section 3.3, one can develop a theory with *multiple* potential threats to peace from different groups, thereby generating conflict "in equilibrium" even in the presence of inefficiencies; see Ray (2010) for some initial steps in this direction. Because a multiplicity of groups are typically formed using ethnic markers, such a theory could also help us understand why ethnic conflict might be salient.[33]

Appendix A

[32] Results are robust to considering other specifications of the restricted model. For instance, since our variables of interest are interactions, we have also used models where the only controls are variables included in the corresponding interaction. That is, the restricted model includes SIZE and OIL or SIZE for the private and public interactions, respectively. The results continue to hold.

[33] Other factors that bear on the salience of ethnic violence includes the greater visibility of ethnicity (Caselli and Coleman, 2013), or the ability of an economically unequal ethnic group to exploit the synergy of money and labor when engaging in conflict (Esteban and Ray, 2008).

Table A.1
Variable Description. Summary statistics for the main variables defined in Appendix A.1.

	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
SHARE CONFLICT	1475	.030	.123	0	.982
SIZE	64001	0.10	0.23	0.00	1.00
OIL	64001	0.00	0.001	0.002	0.01
OIL PC	62103	0.00	0.00	0.00	0.17
MINES	65639	0.57	1.42	0.00	13.00
HOME	61968	0.09	0.20	0.00	1.01
PRIVINDEX	61968	−0.00	0.70	−0.41	5.37
LACK RIGHTS	42950	0.64	0.28	0.00	1.00
CIVIL RIGHTS	42950	0.62	0.27	0.00	1.00
POLITICAL RIGHTS	42950	0.66	0.31	0.00	1.00
CHILD MORTALITY	60669	100.88	50.60	12.54	211.01
AUTOC	45870	0.53	0.29	0.00	1.00
EXCLUDED	63544	0.86	0.34	0.00	1.00
PUBINDEX	38049	0.00	0.96	−2.17	1.32
GIP	64001	0.14	0.35	0.00	1.00
GROUPAREA	64001	84.28	406.74	0.04	7354.72
AREA PC	62103	−9.73	2.00	−15.90	1.55
SOILCONST	64001	1.62	0.78	0.00	6.15
DISTCAP	64001	0.92	1.03	0.00	7.97
MOUNT	64001	0.37	0.36	0.00	1.00
PARTITIONED	64001	0.62	0.48	0.00	1.00
GDP	56945	7.75	1.16	5.08	11.16
POP	61893	17.08	1.81	11.73	20.98
POLITY	58120	−0.09	0.70	−1.00	1.00

A.1. Variable definitions

The following empirical variables are used.

Conflict INCIDENCE: group-level dummy = 1 for those years in which an ethnic group is involved in armed conflict against the state with over 25 battle-related deaths. Source: Cederman et al. (2009); CBR henceforth.

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

SHARE OF CONFLICT YEARS: group-level variable capturing the share of years a group has been in conflict against the State over 1960–2006. Source: CBR.

SHARE OF ONSET YEARS: group-level variable that captures the share of years a group has started in conflict against the State (onset years) in the period 1960–2006. Source: CBR.

SIZE: relative size of the group. Source: CBR.

OIL: log of the product of the homeland area covered by oil (in ’000 km^2) and the international price of oil. To avoid taking the log of zero, 1 has been added to all observations. Source: Oil fields: Petrodata (Lujala and K.Rod, 2007). Oil prices: the World Bank.

OIL CONCENTRATION: Herfindahl index of oil reserve distribution across groups. Source: Petrodata and GREG.

OIL (COUNTRY): log of the area of the country covered by oil (in thousands of square kilometers) 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 from Petrodata (Lujala and K.Rod, 2007). Data on oil prices from the World Bank.

MINES: mineral availability in the ethnic homeland, computed as follows: we consider 13 minerals (bauxite, coal, copper, diamond, gold, iron, lead, nickel, platinum, phosphate, silver, tin and zinc) for which international price data is readily available. For each mineral, year and ethnic group, we create a dummy variable = 1 if the group has at least one active mine of that mineral. Then, each dummy is multiplied by its normalized international price, constructed as the log of its international price divided by the log of its price in 1980 (the year when the data starts). MINES is the sum of the resulting values. Data on mineral availability comes from the *Raw Material Data* dataset, (IntierraR, 2015) whereas data on mineral prices is provided by the World Bank.

PRIVINDEX: Index of privateness computed via factor analysis on the variables OIL, MINES and HOME.

AUTOC: country average of the Polity IV autocracy index for the years 1960 to 1975, normalized between 0 and 1, where 1 is the highest degree of autocracy. Source: Polity IV (2014).

EXCLUDED: average over 1960–1975 of *excluded*, a dummy variable that is 1 if the ethnic group is in power in a given country and year (source: CBR).

CIVIL RIGHTS: Lack of civil liberties from Freedom House. We rescale the original index so it is measured between 0 and 1 (where 0 indicates highest level of civil liberties). For each country, we average its value from 1972 to 1975 and assign the resulting quantity to all post 1975 years. Source: Freedom House (2018).

POLITICAL RIGHTS: Lack of political liberties from Freedom House, rescaled to lie between 0 and 1 (0 = highest level of liberties). For each country, we average its value over 1972–1975 and assign the resulting quantity to all post-1975 years. Source: Freedom House (2018)

LACK RIGHTS: average of CIVIL RIGHTS and POLITICAL RIGHTS.

CHILD MORTALITY: Deaths of children under 5 per 1000 live births. For each country, we consider the average of this quantity over the period 1960 to 1975 and assign it to all subsequent years. Source: UNICEF Global Databases.

PUBINDEX: Index of publicness computed via factor analysis on the variables POLITICAL RIGHTS, CIVIL RIGHTS, AUTOC, EXCLUDED and CHILD MORTALITY.

GIP: dummy = 1 if group is in power in country and year (lagged one year). Source: CBR.

GROUPAREA: area of ethnic homeland (in '000 km^2). Source: GREG.

HOME: area of ethnic homeland relative to total area of country. Source: GREG.

LAND PC: log of the total area divided by group population. Source: GREG.

SOILCONST: measures limitations of homeland soil to agriculture, constructed using the Harmonized World Soil Database from Fischer et al. (2008). Fischer et al. (2008) construct a global grid of 38 nutrient availabilities 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 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 for the group's share of mountainous terrain. Source: CBR.

PEACEYEARS: number of years since the last group-level onset. Source: CBR.

LAG: lagged conflict incidence. Source:CBR.

PARTITIONED: dummy variable = 1 if ethnic homeland covers two or more countries. Source: GREG.

GDP: log GDP per capita, lagged one year. Source: Penn World Tables.

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

POLITY: Polity 2 index, lagged one year. Source: Polity IV (2014).

A.2. Summary statistics for main variables

Provided in Table A.1.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2021.102759.

References

Abadie, A., Athey, G.W., Imbens, J., Wooldridge, S., 2017. When Should You Adjust Standard Errors for Clustering?. Mimeo.

Ai, C.R., Norton, E.C., 2003. Interaction terms in logit and probit models. Econom. Lett. 80, 123–129.

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. J. Polit. Econ. 80, 151–184.

Angrist, J.D., Pischke, J., 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. J. Econ. Perspect. 24, 3–30.

Bannon, I., Collier, P., 2003. Natural Resources and Violent Conflict. The World Bank, Washington D.C..

Bellows, J., Miguel, E., 2009. War and local collective action in Sierra Leone. J. Public Econ. 93, 1144–1157.

Berman, N., Couttenier, D., Rohner, M., Thoenig, M., 2015. This Mine Is Mine! how Minerals Fuel Conflict in Africa. Mimeo.

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

Caselli, F., Coleman, W.J., 2013. On the theory of ethnic conflict. Journal of the European Economic Association 11, 161–192.

Cederman, L.-E., Buhaug, H., Rod, J.K., 2009. Ethno-nationalist dyads: A GIS-based analysis. J. Conflict Resol. 53, 496–525.

Chamberlin, J., 1974. Provision of collective goods as a function of group size. Am. Polit. Sci. Rev. 68, 707–716.

Doyle, M., Sambanis, N., 2006. Making War and Building Peace. Princeton University Press, Princeton.

Dube, O., Vargas, J., 2013. Commodity price shocks and civil conflict: Evidence from Colombia. Rev. Econom. Stud. 80, 1384–1421.

Esteban, J., Mayoral, L., Ray, D., 2012a. Ethnicity and conflict: An empirical investigation. Amer. Econ. Rev. 102, 1310–1342.

Esteban, J., Mayoral, L., Ray, D., 2012b. Ethnicity and conflict: Theory and facts. Science 336, 858–865.

Esteban, J., Ray, D., 1994. On the measurement of polarization. Econometrica 62, 819–851.

Esteban, J., Ray, D., 2001. Collective action and the group size paradox. Am. Polit. Sci. Rev. 95, 663–672.

Esteban, J., Ray, D., 2008. On the salience of ethnic conflict. Amer. Econ. Rev. 98, 2185–2202.

Esteban, J., Ray, D., 2011. Linking conflict to inequality and polarization. Amer. Econ. Rev. 101, 1345–1374.

Fearon, J., 2003. Ethnic and cultural diversity by country. J. Econ. Growth 8, 195–222.

Fearon, J., 2005. Primary commodity exports and civil war. J. Conflict Resol. 49, 483–507.

Fearon, J., Laitin, D., 2003. Ethnicity, insurgency, and civil war. Am. Polit. Sci. Rev. 97, 75–90.

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

Freedom House, 2018. Freedom of the world, dataset. https://freedomhouse.org/report/freedom-world-aggregate-and-subcategory-scores.

Gleditsch, N.P., Wallensteen, M., Eriksson, M., Sollenberg, P., Strand, H., 2002. Armed conflict 1946–2001: A new dataset. J. Peace Res. 39, 615–637.

IntierraR, M.G., 2015. SNL metals and mining, dataset. http://www.snl.com/Sectors/metalsmining/Default.aspx.

Le Billon, P., 2001. The political ecology of war: Natural resources and armed conflicts. Polit. Geogr. 20, 561–584.

Licklider, R., 1995. The consequences of negotiated settlements in civil wars, 1945–1993. Am. Polit. Sci. Rev. 89, 681–690.

Lujala, P., 2010. The spoils of nature: Armed civil conflict and rebel access to natural resources. J. Peace Res. 47, 15–28.

Lujala, P., K.Rod, J., 2007. Fighting over oil: Introducing a new dataset. Conflict Manage. Peace Sci. 24, 239–256.

Marwell, G., Oliver, P., 1993. The Critical Mass in Collective Action. a Micro-Social Theory. Cambridge University Press, Cambridge, UK.

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.

Mill, J.S., 1859. On Liberty. Parker and Son, London.

Morelli, M., Rohner, D., 2015. Resource concentration and civil wars. J. Dev. Econ. 117, 32–47.

Oliver, P., Marwell, G., 1988. The paradox of group size in collective action: A theory of the critical mass. II. Am. Sociol. Rev. 53, 1–8.

Olson, M., 1965. The Logic of Collective Action. Harvard University Press, Cambridge, MA.

Oster, E., 2019. Unobservable selection and coefficient stability: Theory and evidence. J. Bus. Econom. Statist. 37, 187–204.

Pareto, V., 1927. Manual of Political Economy. A.M. Kelley, New York.

Polity IV, 2014. Political regime characteristics and transitions, dataset. http://www.systemicpeace.org/inscrdata.html.

Ray, D., 2010. Multiple Threats To Peace. mimeo. Department of Economics, New York University, https://debrajray.com/wp-content/uploads/2021/08/RayMultipleThreats.pdf.

Sambanis, N., 2001. Do ethnic and non-ethnic civil wars have the same causes? A theoretical and empirical enquiry. J. Conflict Resol. 45, 259–282.

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

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

Skaperdas, S., 1996. Contest success functions. Econom. Theory 7, 283–290.

Taylor, M., 1987. The Possibility of Cooperation. Cambridge University Press, Cambridge, UK.

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

Tullock, G., 1980. Efficient rent seeking. In: Buchanan, J., Tollison, R., Tullock, G. (Eds.), Toward a Theory of the Rent-Seeking Society. Texas A & M University Press, College Station, TX, pp. 97–112.

Weidmann, N.B., Rod, J.K., Cederman, L.-E., 2010. Representing ethnic groups in space: A new dataset. J. Peace Res. 47, 491–499.