

Identity conflict, ethnocentrism and social cohesion

Matteo Sestito

WP 2023 Nr 04

Identity conflict, ethnocentrism and social cohesion*

Matteo Sestito[†]

January 23, 2023

Abstract

This paper uses a novel dataset on ethnic warfare to shed light on how conflict affects social identification and cohesion. A large body of anecdotal studies suggests that ethnic identities become more salient at times of conflict. Using data from eighteen sub-Saharan countries, I provide econometric evidence for such a claim. The effect of ethnic conflict on various measures of social cohesion is also investigated, uncovering a positive relationship between the two. The finding is understood as a result of the ethnocentric dynamics generated by conflict: as ethnic warfare increases ethnic identification, in-group cooperation follows suit. This parochial interpretation is further strengthened by the use of remote violence and the conditionality of conflict-induced pro-social behaviour on low levels of ethnic polarisation.

JEL codes: D74, N47, O55, Z13

Keywords: Ethnic conflict, social cohesion, identity, Africa

*I would like to thank my advisor, Nicolas Berman, for all the insightful advises. Comments by Cecilia Garcia Peñalosa, Garance Genicot, Pauline Grosjean, and Yannick Dupraz have been of great help. This paper has benefited from useful comments by seminar participants at Aix-Marseille School of Economics, Pompeu Fabra University, 2021 Annual Southern PhD Economics Conference, 2021 International Conference on Development Economics, 2021 Annual Workshop of the Households in Conflict Network, 2021 Annual meeting of the Association of Southern European Economic Theorists, 2022 Louis-André Gérard-Varet conference in public economics. It goes without saying that any error in the work is my own responsibility. This work was supported by the French National Research Agency Grant ANR-17-EURE-0020 and by the Excellence Initiative of Aix-Marseille University - A*MIDEX.

[†]Aix-Marseille University, CNRS, EHESS, Centrale Marseille, AMSE, France.

1 Introduction

Now, it is time to bring in the harvest

Hutu radio

Frightening messages of this kind were aired by the Hutu propaganda throughout the 1994 genocide: in less than 100 days, over half million of cadavers dotted the green slopes of Rwanda.¹ The Hutu government had launched an all-out war against the country's Tutsi minority, with civilians routinely slaughtered because of their ethnic affiliation. External observers remained baffled at the ferocity and violence of the conflict, which was described as a war of all against all, of neighbour against neighbour. People sought shelter into their ethnic community, turning against those hailing from other cultural backgrounds. While journalists and politicians alike have struggled to make sense of these dynamics, invoking ill-defined stereotypes and 'ancient hatreds' (Allen and Seaton 1999), the academia has since long recognised the role that instrumental considerations have had in shaping these culturally-oriented behaviours of cooperation and discrimination. Demystifying the notion of an Hobbesian hell, one observer of the Rwandan conflict concluded that: "the chief dynamic of the depredations seems to have been furnished by marauding bands of violent, opportunistic, and often drunken thugs" (Mueller 2000, p. 62).

The discussion surrounding the Rwandan genocide is symptomatic of a broader debate on the saliency and importance of social identities at times of conflict: are ethnocentric pulses a real by-product of conflict or are they a mere facade used instrumentally to justify opportunistic conduct? The question is important not only to analyse conflict dynamics, but also to comprehend post-war social norms and institutions, with crucial repercussions for policy-making.² This paper takes part to the debate by analysing the interplay between warfare and ethnocentrism, and exploring its consequences in terms of cooperative behaviour. The topic has, indeed, received growing attention in the last couple of decades and an increasing number of studies has explored the impact of conflict on pro-social attitudes. While establishing a positive link between the

1. For the quote see Peterson (2000, p. 273).

2. For example, the debate is of paramount importance for the delineation of settlement patterns in post-conflict environments. On the benefits and costs of ethnic segregation in the aftermath of sectarian wars see: Kaufmann (1996), Laitin (2004), Mironova and Whitt (2016).

two, evidence on the actual channels is scant and “theoretical mechanisms remain elusive” (Mironova and Whitt 2021, p. 2). The present paper fills this gap by exploring the effect of ethnic conflict on social identification and cohesion across sub-Saharan Africa. Thanks to a newly created dataset on sectarian warfare, I provide empirical support to the idea that conflict triggers a genuine process of identity radicalisation and that these ethnocentric feelings are the main channel whereby conflict translates into deeper cooperation.

Large scale quantitative evidence on the subject is, indeed, still very limited. A major barrier to empirical investigation in the field is the lack of comprehensive data on ethnic violence, that is conflict activity (allegedly) carried out for the sake of a given ethnic group.³ To overcome such limitation, I compiled a novel dataset. Using geo-coded data from the Armed Conflict Location and Event Data (ACLED) Project, I manually match 1082 armed actors to various sub-Saharan ethnolinguistic groups. The process allows the construction of conflict measures that vary across time, space, and ethnicity. These variables are then connected to several Afrobarometer surveys, describing social attitudes and identification across the African continent.⁴ In particular, I employ three measures of social cohesion aimed at capturing pro-social cooperative behaviour at the local level. These are: attendance at community meetings, membership in community development associations, and participation in instances of collective action. As for social identity, Afrobarometer asks respondents whether they feel more represented by their ethnic group or by their nation-state.

The triple source of variation of the conflict variables is the essence of the identification strategy here employed. The baseline model consists of an Ordinary Least Squares (OLS) regression of social behaviour on ethnic conflict, as enriched by cell-district \times time and ethnic \times time fixed effects, other than various individual-level covariates taken from the Afrobarometer. Hence, the baseline econometric specification compares people exposed to the *same* conflict incident, but with different treatment status because of their ethnolinguistic affiliation. While, obviously, violence does not unfold randomly, the rich set of fixed effects attenuates issues of reverse causality (*e.g.* ethnicities self-selecting into conflict) and

3. A notable exception is the dataset developed by Guarnieri and Tur-Prats (2022). Their data, however, refer exclusively to major armed conflicts.

4. The choice of sub-Saharan Africa as a setting for the analysis responds primarily to issues of data availability. Moreover, ethnic conflict is concentrated in sub-Saharan Africa (Fearon and Laitin 2003), where sub-national identities are still strong and considerably affect the course of public policy (Easterly and Levine 1997).

problems related to omitted variables (*e.g.* different local development trends). It remains, however, the possibility that estimates are biased by unobservable confounding factors operating at the ethnolinguistic-geographic-temporal level. A series of robustness tests aims at lowering these concerns, and gives credit to an interpretation of the results beyond the mere correlation. Among the various sensitivity checks, the most relevant is the use of *distant* warfare, that is, ethnic conflict taking place outside the cell-district of each observation. Conflict exposure is therefore indirect and likely orthogonal to local (ethnic-specific) shocks affecting social norms.

The models here estimated substantiate a positive impact of ethnic conflict on in-group identification and cohesion. First, ethnic conflict makes ethnic identity more salient. The result echoes popular perceptions and gives empirical support – the first on a cross-country level – to the large qualitative literature linking warfare to episodes of ethnic radicalisation.⁵ Moreover, the finding qualifies those studies reducing identity shifts to mere opportunistic calculations (Kalyvas 2008, Mueller 2000).

The second major finding of the paper is that ethnic conflict has also a positive effect on the three selected measures of social cohesion. The result is in line with some recent works relating warfare to social capital (Bauer et al. 2016). Importantly, the research design here employed allows to test a plausible mechanism behind this positive link. Warfare, when externally felt by a group, can, indeed, increase pro-social behaviour through several channels. These, for the purposes of this paper, can be divided into two non-exclusive categories.

First, warfare can change social norms by providing direct *material* incentives favouring pro-social strategies. These incentives stem from both economic and security rationales. On the one hand, amidst the insecure environment generated by conflict, communal insurance mechanisms acquire greater importance, possibly paving the way to stable cooperative institutions and norms (Gilligan et al. 2014). On the other hand, conflict triggers a demand for security that is, indeed, often met

5. See, for example, Kaufmann (1996, pp. 140-145) for an overview of how ethnic identification hardened throughout many 20th century conflicts. Two of the most influential case studies in the literature are the 1994 Rwandan genocide and the ethnic violence following the collapse of Yugoslavia in the early 1990s. For the first the reader is referred to Peterson (2000). On the second, the excellent anthropological documentary filmed by Bringa and Christie (1993) remains an unrivalled window into the conflict; for a more general overview of the Yugoslavian wars see Sambanis and Shayo (2013).

by the creation of local self-defence forces (Jennings and Sanchez-Pages 2017). Cooperation within the latter can then gradually extend to the community as a whole.⁶

The second category of mechanisms linking conflict to cooperation, works through psychological processes that directly change individuals' social preferences. One of these channels operates through ethnocentric dynamics: conflict results into stronger affiliation towards co-ethnics, ultimately easing cooperation within one's own group. The very fact that warfare unfolds along ethnic cleavages makes ethnic attributes more salient: individuals perceive themselves into ethnolinguistic terms and the group identity is made thicker by the war experience.⁷ Hence, as ethnocentric feelings are prioritised, people become more willing to cooperate with their co-ethnics, who are felt closer and are equally sympathetic to those violent clashes investing their ethnolinguistic group as a whole.

There are, however, many other psychological processes linking warfare to cooperation. These mechanisms range from spontaneous outrage against war injustices – with the related effort to redress them – to theories explaining public engagement as an emotional value (Bateson 2012). The one which has received the most attention is, perhaps, the so-called post-traumatic growth theory. According to the latter, in the aftermath of traumatic events people recast their beliefs, often in a drastic manner. The reformation of one's own *weltanschauung* goes in the direction of a renewed appreciation of one's own: capability, spiritual dimension, and intimate relations (Tedeschi and Calhoun 2004); thus possibly leading to higher involvement in social and public life.⁸

6. There is some evidence on the emergence of local self-defence forces within the period here considered. This is the case, for example, of the 'Moorland Forces Militia (Kenya)', a Mosop self-defence group born to counter the Soi-led 'SLDF: Sabaot Land Defence Force' active in Mount Elgon (Kenya) (Simiyu 2008).

7. Conflict is, indeed, a great repertory of images and symbols shaping and substantiating the social identity and cohesion of a group. Beyond the mythologisation of battles as defining moment of the identity-building process – the centuries-long idealisation of the battle of Poitiers (732 CE) as defining moment for the definition of a European identity, despite its marginal strategic importance at the time, is a case point –, victimisation and, possibly, revenge appear to be of paramount importance in cementing a common sense of identity. As an outstanding contemporary example, consider Israel, where Holocaust martyrs are revered in the poignant Yad Vashem museum of Jerusalem and the related sense of victimisation is still today a crucial element in the definition of Israeli national feelings (Resnik 2003).

8. Explicit empirical evidence of post-traumatic growth with respect to conflict remains, however, quite elusive. For some limited evidence concerning Yugoslavia and Liberia, see Powell et al. (2003) and Hartman and Morse (2020), respectively.

While these mechanisms may coexist and reinforce one another, the observed increase in social cohesion seems to be driven by the ethnocentric dynamics inherent to sectarian conflict. First, the use of ethnic-level conflict implies a cultural, rather than geographical, link between warfare and individual behaviour. This, also in light of the positive impact on ethnic identification, suggests a parochial interpretation of the results: the observed increase in pro-social behaviour remains limited to one's own co-ethnics. The intuition is further tested by checking for the impact of distant conflict activity, whose effect is unlikely to operate through institutional channels affecting local economic and security incentives. Moreover, when (local) ethnic and non-ethnic violence are compared, only the first has a positive significant impact on pro-social behaviour. Finally, the models estimated at Section 5 illustrate how the effect of ethnic warfare is heterogeneous with respect to local measures of ethnic polarisation: conflict-induced cooperation fails to materialise in communities highly polarised across ethnic lines, further giving strength to a parochial understanding of the baseline results.

Overall, the analysis provides empirical evidence to the idea that ethnic boundaries have a pivotal role in shaping the content and extent of pro-social behaviour during periods of conflict. While instrumental considerations have been correctly identified as important factors governing group cooperation, identity-based processes, such as those underlined here, are equally relevant in explaining post-conflict social norms.

The rest of the paper is organised as follows. The next section reviews the existing evidence and the related literature. Section 3 describes the data used in the empirical analysis, which is then presented in Section 4. Hence, I proceed to test the ethnocentric mechanism as explanation for the conflict-induced increase in social cohesion. A final section concludes.

2 Existing evidence

Given the limited availability of data on ethnic conflict, few studies have quantitatively explored its consequences in terms of social identification.⁹ The most notable exception is the work by Rohner and co-authors (2013), who investigate the issue in relation to the resurgence of the Lord's Resistance Army insurgency in Uganda during the early 2000s. While their main model uses warfare on a geographical base, in an exten-

9. For a theoretical treatment of the topic see Sambanis and Shayo (2013).

sion they match conflict to ethnolinguistic groups. They find a positive impact of ethnic conflict on ethnic identification, which lead them to a conclusion similar to that of this paper: conflict impact on inter-personal trust is mediated by ethnocentric dynamics. Beyond the obvious gains in terms of external validity, this paper employs a more demanding empirical specification than that used by Rohner et al. (2013), which does not include spatio-temporal fixed effects.

Other papers analyse the responsiveness of social identity to violence, but using geographically-defined conflict and thus focusing on identity shifts due to the changing balance of power across groups. Ananyev and Poyker (2021), for example, provides another case study from sub-Saharan Africa. The authors look at the Tuareg insurgency in Mali (2012), finding that conflict-induced lower perception of state capacity is associated to decreases in national identification. The only cross-country work is that of Besley and Reynal-Querol (2014), who show that historical conflict in Africa is correlated to higher ethnic identification and more inter-ethnic mistrust. In departure from this literature, the present work deals with ethnolinguistic warfare. Thanks to the newly developed dataset linking conflict actors to ethnic groups, I can specify a cultural relationship between warfare and ethnic identity. This research design can control for local institutional dynamics, hence distinguishing changes in the saliency of ethnic affiliation from shifts in national identity.

A second body of literature to which this paper is related concerns those articles analysing the impact of conflict on pro-social behaviour. Interest in the field, at least among economists, was sparked by Bellow and Miguel's (2009) study in Sierra Leone. Using a household survey reporting also for exposure to the 1991-2002 civil war, they found that people more exposed to conflict were also more likely to participate in community life once hostilities were over. Similar findings have been echoed by studies of various other conflicts around the globe, such as those in: Burundi (Voors and Bulte 2014), Uganda (De Luca and Verpoorten 2015a-b), Russia (Guriev and Melnikov 2016), and Nepal (Gilligan et al. 2014). Moving beyond the case-study literature, Grosjean (2014) analyses the legacy of violent conflict in thirty-five Eurasian countries. She finds that personal or familiar exposure to warfare is associated to higher levels of social capital. In a similar vein, using data at the global level, Bateson (2012) concludes that crime victimisation positively affects attendance at community meetings. The findings of the literature are not, however, completely uncontroversial. Conflict

type and degree of exposure seem, indeed, to matter a lot. Blattman (2009) considers the consequences of abduction by the Lord's Resistance Army during the Ugandan civil war: in terms of pro-social behaviour, children who were captured and forced to fight with rebels are indistinguishable from other civilians. Kibris and Cesur (2022) study pro-social behaviour among Turkish conscripts engaged in the counter-insurgency aimed against the country's Kurdish minority. Exploiting quasi-experimental variation, they show that exposure to armed conflict does not predict community participation among former conscripts. Cassar et al. (2013) analyse the internal conflict in Tajikistan (1992-1997), finding a negative relationship between victimisation and within-community trust. The authors provides empirical support to the intuition that their results are explained by the intra-community nature of the Tajik war.

Overall, one of the conclusions that can be drawn from this literature is that *external* warfare can induce cooperation, but only limited to the group that is swamped by conflict as a 'unitary' actor.¹⁰ When conflict, instead, cuts through internal cleavages, social capital tends to deplete because, as a consequence of the insecure environment, mistrusting attitudes become prioritised and social networks collapse. The focus on ethnic conflict of this paper allows a sharp definition of the groups involved in conflict, thus providing clear evidence of the in-group bias of conflict-induced cooperation.¹¹ More importantly, different mechanisms can be here compared, closing a major gap in the literature. The use of distant ethnic violence and the other tests carried out at Section 5, unambiguously identify ethnocentrism as the mechanism underlying the relationship between conflict and social cohesion.

10. Mironova and Whitt (2018 & 2021) provide explicit evidence on the parochial nature of the pro-social impact of warfare. In a series of laboratory experiments, the authors show that in the aftermath of conflicts in Kosovo (1998-1999) and Ukraine (2014-ongoing), participants in various cooperative games display a strong positive bias towards co-ethnics.

11. This parochial interpretation, although not always explicitly spelt out, is also coherent with those articles finding a negative impact of conflict on: generalised trust (Besley and Reynal-Querol 2014, Rohner et al. 2013), central political institutions (De Juan and Pierskalla 2016, Grosjean 2014, Voors and Bulte 2014), and judicial impartiality (Shayo and Zussman 2017). It shall be noted that some of these results have been contradicted in other studies, identifying a positive effect of conflict on impersonal trust (Bellows and Miguel 2009, De Luca and Verpoorten 2015b) and participation in public politics (Bateson 2012, Bellows and Miguel 2006 & 2009, Blattman 2009, De Luca and Verpoorten 2015a, Gilligan et al. 2014). A review of the reasons for these discrepancies is out of the scope of this paper.

A final strand of literature related to this paper concerns the inter-play between state-building and ethnicity in sub-Saharan Africa. Most of the countries in the continent are highly multi-ethnic; for example, the sole Nigeria is credited with more than 250 different ethnolinguistic groups. While the co-existence of different peoples shall not necessarily hamper political stability, in the African context, marked by historically weak institutions (Herbst 2000, Acemoglu and Robinson 2012, Hopkins 2019), multiculturalism has often been a source of troubles. Many states are, indeed, characterised by fierce political competition along ethnic lines. The contest often assumes a ‘winner-takes-all’ connotation, whereby the group in power squeezes public resources as much as possible for the sake of its fellow co-ethnics. Some of the most blatant examples include: early post-independence Ghana (Easterly and Levine 1997), Nigeria at least until 1999 (Kesselman et al. 2016, ch. 12), and Kenya since the Moi’s regime (Elischer 2013, ch. 3). Moreover, given also the dis-functionality of central administrations, independence claims have abounded in the post-colonial period. Only two of them have ultimately proved successful: Eritrea gained independence from Ethiopia in 1991, and South Sudan separated from Sudan in 2011. Many more attempts have been suffocated in blood, leaving open scars in the collective memory of several peoples. Hence: the cycles of Tuareg rebellions in Mali, the Igbo struggle in Nigeria, the endless independence conflict in the Casamance region of Senegal, etc. The positive link between sectarian violence and ethnic-vs-national identification here found, unveils one of the mechanisms for the persistence of weak central administrations in sub-Saharan Africa.

3 Data

A Sample and outcome variables

Outcome variables on ethnic identity and social cohesion are taken from Afrobarometer surveys. These are nationally representative surveys routinely used in the economic literature (*e.g.* Nunn and Wantchekon 2011, Michalopoulos and Papaioannou 2013). Here, I employ rounds 3 to 6, which cover the ten year window from March 2005 to September 2015. I keep only the eighteen countries that are present in all the surveys.¹²

12. In principle, the research design allows for the inclusion of any country appearing in at least two surveys. There are, indeed, other sixteen countries satisfying such criterium. Their exclusion is based on reasons of interpretability and time, with any

These are: Benin, Botswana, Cape Verde, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Namibia, Nigeria, Senegal, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe. They are depicted in Figure 1, which also reports their subdivision in cells of 0.5×0.5 degrees ($\sim 55\text{km} \times 55\text{km}$).¹³

Crucially, Afrobarometer provides information on the language of respondents, allowing their classification in ethnolinguistic groups. I use the LEDA algorithm developed by Müller-Crepon and co-authors (2020) to harmonise languages across Afrobarometer rounds.¹⁴ The procedure yields a sample that comprehends 111,428 individuals, belonging to 247 ethnolinguistic groups and spread across 1938 cells in eighteen countries.¹⁵

The question capturing social identification is expressed in terms of national-vs-ethnic allegiance. It varies discretely from one (the respondent feels only national) to five (the respondent identifies fully with her ethnic group). Note that this variable has been inverted around its median for easing the interpretation of results. The original coding of some other Afrobarometer controls has also been changed in order to harmonise them across survey rounds. The exact construction and definition of all variables, alongside their recoding, is given in Appendix A. In robustness exercises, some variables measuring for political trust are used for ensuring that shifts in social identification are due to movements in ethnic, rather than national, allegiance.

The variables describing social cohesion are: attendance at community meetings, membership in community associations, and involvement in instances of collective action. They are standard measures of pro-social behaviour at the community-level, abundantly used in the literature relating conflict to the latter. They are expressed in integers, with higher values signalling greater involvement in the local community. For clarity of exposition, I standardise the three measures and then average

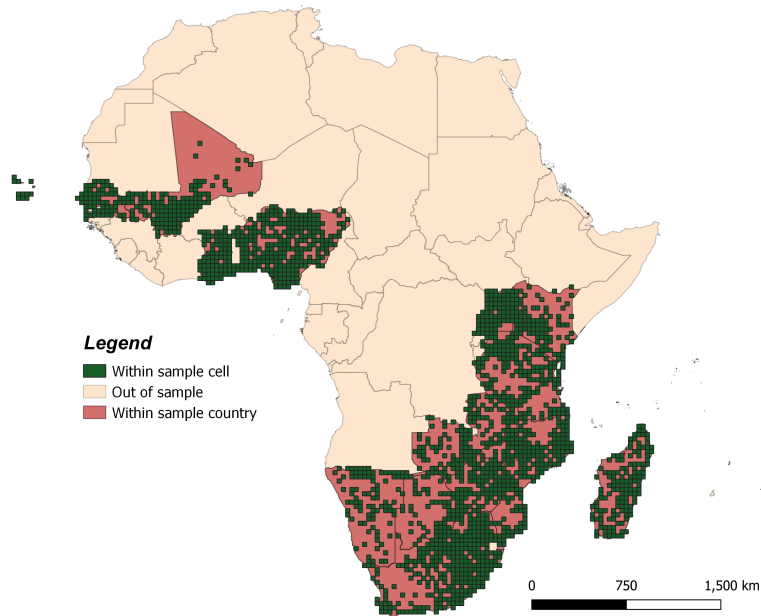
new country included in the analysis considerably increasing the number of armed groups necessitating a match to an ethnicity.

13. This is the standard size of grid cells typically used in the conflict literature. See for example: Berman et al. (2017), Manacorda and Tesei (2020).

14. Few languages could not be matched using this procedure; they have been excluded from the analysis.

15. Note that Afrobarometer also reports the ethnic group of the interviewed people. The use of language to define ethnolinguistic groups is dictated by the fact that there are 8548 less missing observations and that LEDA algorithm is based on the Ethnologue linguistic tree. The practice is also standard in the literature (*e.g.* Depetris-Chauvin et al. 2020). As reported in the paper and the appendix, results are robust to using ethnicity as an ethnolinguistic marker.

Figure 1: Countries and cells in Afrobarometer surveys 3 to 6



them into a social cohesion index. In Appendix D I show that results hold also for each individual outcome. Table 1 illustrates their descriptive statistics.

B Ethnic conflict

Data on conflict is taken from the ACLED database, which has been widely used in economics.¹⁶ It covers conflict incidents spanning the whole African continent since 1997, providing information on their location, date, and actors involved.¹⁷ In order to define ethnic conflict, I analyse all the non-state armed groups appearing in the within-sample countries and in their neighbourhood during the relevant period. There are 1082 of them in total. I associate a conflict actor to an ethnolinguistic group when the former is organised along ethnic lines. Matches

16. To cite just some recent work: Depetris-Chauvin et al. (2020), Manacorda and Tesei (2020), Moscona et al. (2020).

17. The wide coverage of ACLED, allowing also for the inclusion of low-scale localised violence, is the chief reason why I employ this dataset in lieu of other geo-localised databases, most notably the UCDP-GED dataset. The latter includes only deadly conflict incidents belonging to a dyad crossing at least once the 25 battle-related deaths per year.

have been based on a *wide* variety of sources: primarily scholarly publications, policy reports and international news agencies, but also, in the case of minor ephemeral actors, on social media posts and local newspapers. The ethnolinguistic dictionary of Olson (1996) has been the main reference used to harmonise ethnic affiliations towards the Afrobarometer categories. Details on the sources used as well as the rationale behind each actor-ethnicity link, are reported in the dataset and in its codebook. Appendix B gives a more accurate overview of the database.

A typical positive match is for example that between the ‘Arrow Boys Militias’ and the ‘Ateso’ ethnolinguistic group. The militia was a local self-defence force in the Teso region of Uganda, which is inhabited by the Ateso-speaking Iteso people. Its formation was incentivised by the government to fight off the Lord’s Resistance Army (LRA). It had a uniquely regional character and framed the fight against the LRA in ethnic terms.¹⁸

Not all matches were so uncontroversial. Armed groups have often multiple identities and objectives, so that their ethnic dimension is sometimes difficult to ascertain. These cases have been explicitly coded and they are dropped in robustness exercises. An example of uncertain match is that between the ‘NDFF: Niger Delta Freedom Fighters’ and the ‘Ijaw’ ethnolinguistic group. The actor has a popular base among Ijaw people, with its membership and symbolism referring mostly to the Ijaw world. Yet, its leadership is controlled by the Urhobo ethnic group (Oriola 2013, pp. 106, 146-147).

Another problem encountered during the matching procedure was that Afrobarometer languages can sometimes encompass the ethnolinguistic group represented by the ACLED actor. As example of the latter, consider the ‘Ilajes Ethnic Militia (Nigeria)’. The Ilajes are a sub-group of the Yoruba people (Olson 1996, p. 239), which is the ethnic category reported in the Afrobarometer surveys. There is a fair chance that not all Yorubas are deeply concerned about Ilajes’ fortunes. To account for this, an alternative matching rule excludes all the cases whereby Afrobarometer languages engulf actor’s ethnolinguistic affiliation.

Once conflict is associated to ethnicities, it is aggregated over time and space. Define $c_{e,l,j}$ as a dummy indicating the presence of a conflict event taking place in location l during day j , and involving an armed

18. This match is based on: Lomo and Hovil (2004), the data kindly shared by Rohner et al. (2013), and the online description of the actor provided by the Uppsala Conflict Data Program at: <https://ucdp.uu.se/actor/585>.

actor associated to ethnolinguistic group e .¹⁹ I exclude conflict incidents involving state forces to ease the interpretation of results and to avoid endogeneity problems related to changes in state capacity and legitimacy.²⁰ Further define $ab_{k,t}$ as the earliest interview date conducted by Afrobarometer round $t \in \{3, 4, 5, 6\}$ in cell k . Hence, conflict is aggregated: temporally, over the period between two consecutive Afrobarometer surveys, that is, over the day-unit intervals $AB_{k,t} = \{ab_{k,t-1}, \dots, ab_{k,t} - 1\}$; spatially, over all the locations belonging to the cell k . Formally:

$$\tilde{C}_{e,k,t} = \sum_{l \in k} \sum_{j \in AB_{k,t}} c_{e,l,j}$$

The baseline measure of conflict is an indicator function equal to one if there was at least a conflict incident $c_{e,l,j}$ in the relevant temporal-spatial-ethnic window, that is:

$$C_{e,k,t} = \mathbb{1}[\tilde{C}_{e,k,t} > 0]$$

The event-based definition of conflict (*i.e.* $\tilde{C}_{e,k,t}$) is used, instead, in robustness checks. Another measure used for sensitivity analysis is the count of months with at least a conflict incident.

C Descriptive evidence

Descriptive statistics for the main outcome and treatment variables are reported in Table 1. As shown in the first column, ethnic identification is on average relatively high: its mean (2.36) is slightly above the value indicating a preference of national over ethnic attributes (2). Summary statistics for ethnic conflict are in the last column. This is a relatively rare event, with slightly more than 4% of the sample belonging to an ethnic group involved in a conflict in the relevant spatial-temporal window. The majority of warfare in sub-Saharan Africa does not, indeed, unfold along ethnic lines: while 37% of the cells experienced at least a conflict incident, only 12% of the total witnessed ethnic violence.

19. Note that some of these conflict incidents involved armed actors with the same ethnolinguistic background. Intra-ethnic conflict – which is experienced by only a small fraction (0.8%) of the sample – cuts through the group and thus is potentially unrelated to the ethnocentric dynamics here analysed. In Appendix D I show that, indeed, when excluding intra-ethnic conflict, the estimated marginal impacts are generally larger.

20. In Appendix D it is shown that results are nearly identical when including clashes between ethnic actors and national armies. Note that, anyway, the baseline definition of conflict does not exclude violent events pitting civilians & rioters against armed groups with an ethnolinguistic affiliation.

Table 1: Descriptive statistics

	Ethnic Identity	Social Cohesion	Attendance Community Meetings	Collective Action	Membership Community Associations	Ethnic conflict
Mean	2.36	0.00	2.15	1.86	0.61	0.043
Sd	1.20	0.79	1.30	1.29	0.93	0.203
Min	1	-1.25	0	0	0	0
Max	5	1.89	4	4	3	1
<i>N</i>	103656	109426	110772	110520	110387	111428

Notes: The sample includes Afrobarometer rounds 3 to 6. *Ethnic Identity* captures the saliency of ethnic identification, while *Social Cohesion* is an index combining three dimensions of pro-sociality: attendance at community meetings (*Attendance Community Meetings*), participation in instances of collective action (*Collective Action*), and membership in community associations (*Membership Community Associations*). The *Ethnic conflict* dummy signals the presence of conflict events involving the ethnolinguistic group of an observation and taking place in her cell of residence during the relevant period.

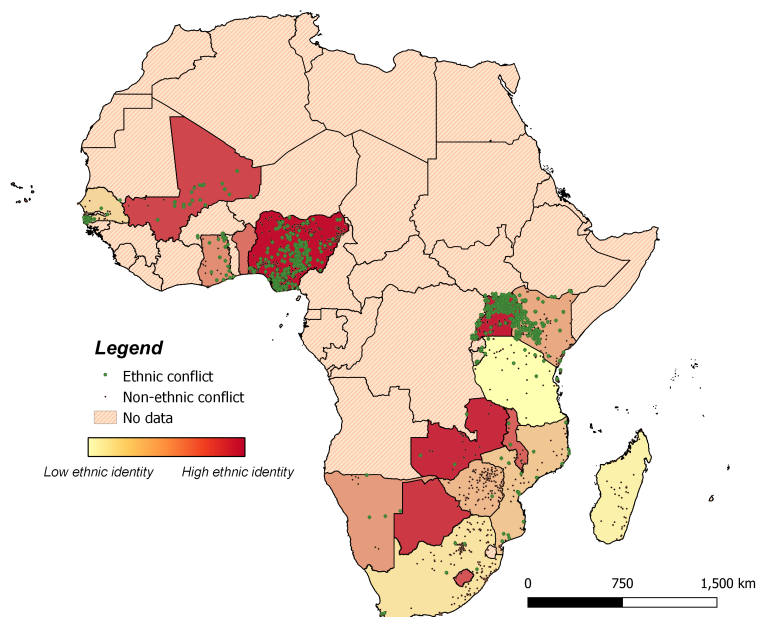
The relative paucity of ethnic violence masks, however, a great variety of conflict types present in the sample. The following are anecdotal observations based on the study done during the match procedure. While not having the pretence of providing a complete description of conflict in sub-Saharan Africa – if anything because clashes between state forces and civilians were not analysed – they are, nonetheless, a rare occasion in which the black box of ACLED conflict is actually open.

The first remark is that there are few ‘wars’, at least in the meaning popularly attached to the word. Conflicts aiming at regime change or territorial conquest, are extremely rare. Examples of ‘wars’ unfolding in the period considered are: the violent independence pulses erupting in the Niger Delta (Nigeria) since the early 2000s; the Tuareg-cum-Islamist rebellion in northern Mali resurfacing around 2012; and al-Shabaab’s insurgency in Kenya (early 2010s-present), which was however largely devoid of an ethnic dimension.

Particularly present in Nigeria and West Africa at large, is conflict between nomadic Fulani herders and various settled farmer communities. Similarly, pastoralists and agriculturalists have repeatedly clashed in Uganda, where Karamajong have often trespassed private fields. More generally, land conflict – in a continent undergoing a demographic transition – is rather widespread across Africa. The intense ethnic clashes between Mosop and Soi communities of Mount Elgon (Kenya), are among the clearest examples.

Finally, within the sample period, there has been a substantial amount of ethnically-organised political violence. Sometimes, this type of con-

Figure 2: Ethnic identity and conflict in selected sub-Saharan countries



flict engulfs the whole state apparatus. This is the case of Kenya, where elections are ferocious contests between ethnic-based parties. In other instances, violence unfolds over local power dynamics, most notably, disputes relative to royal succession in traditional sub-national kingdoms; the Dagbani chieftaincy crisis in Ghana being a case point.

Figure 2 gives an overview of the geographical distribution of ethnic conflict. The map also shades countries in base of their average social identification value. There are some states, such as Botswana, where ethnic identification is relatively strong even in the absence of abundant conflict. Yet, overall, ethnolinguistic conflict tends to be concentrated in countries where ethnic animosities run higher, Nigeria and Uganda, but also Mali, being the clearest examples.²¹ Obviously, the map is at most suggestive, giving a static between-country comparison that masks much sub-national and cross-temporal variation.

21. In Appendix D, I show that results are robust to singularly excluding each country.

4 Econometric analysis

A Baseline model

The baseline econometric model consists of a simple linear equation estimated through OLS:

$$y_{i,e,k,t} = \mu_{k,t} + \lambda_{e,t} + \beta C_{e,k,t} + \delta' x_{i,e,k,t} + u_{i,e,k,t} \quad (1)$$

Where, $y_{i,e,k,t}$ measures social behaviour of individual i belonging to ethnolinguistic group e and interviewed by Afrobarometer round t in cell k . The $\mu_{k,t}$ and $\lambda_{e,t}$ terms are sets of cell-period and ethnicity-period fixed effects.²² The independent variable of interest is the ethnic conflict dummy $C_{e,k,t}$, equal to 1 when ethnolinguistic group e experiences conflict during t within cell k . The vector $x_{i,e,k,t}$ controls for various individual-level characteristics reported in Afrobarometer surveys. In the baseline estimations, it includes only ‘fixed’ personal attributes that are unlikely to be endogenous in equation (1). These are: age and its square, gender, and an urban-rural indicator. Finally, $u_{i,e,k,t}$ are standard errors clustered at the cell-ethnolinguistic-period level, that is, the level of assignment of the conflict variable.²³

B Identification

The identification strategy behind the model exploits the triple source of variation of the ethnic conflict variable. Its coefficient (β) captures the difference in social attitudes between individuals living in the same area and interviewed at the same time, but with different conflict exposure due to their different ethnic background.²⁴ The inclusion of cell \times period and ethnicity \times period fixed effects, means that we are using variation within these groups to estimate the marginal impact of interest. While making the model relatively saturated, the rich set of fixed effects allows to control for many ethnic-specific, possibly time-varying, factors, which

22. Note that while there are two sets of fixed effects, this setting differs from the standard two-way fixed effect (TWFE) model in that there are not fixed groups whose treatment status changes across time. Hence, TWFE diagnostic and remedies (De Chaisemartin and d’Haultfoeuille 2020, Callaway and Sant’Anna 2021), are here irrelevant.

23. In Appendix D it is shown that results are robust to the use of higher-level clustering.

24. Alternatively, the coefficient can be interpreted as a comparison between people belonging to the same ethnic cluster and interviewed at the same time, but having a different exposure to conflict in virtue of their residence.

have been shown to correlate with trust attitudes, conflict propensity, and socio-economic outcomes.²⁵ At the same time, through the $\mu_{k,t}$ term, equation (1) controls for institutional elements likely to influence both conflict and social cohesion: land tenure regimes (Boone and Nyeme 2015), national sport competitions (Depetris-Chauvin et al. 2020), elections (Posner 2004, Eifert et al. 2010), and the overall degree of modernisation and state-building (Miguel 2004, Robinson 2014).

Migration is another obvious confounding factor: in 2016, across sub-Saharan Africa, there were more than 12 million internally displaced people for reasons related to conflict and violence.²⁶ These migratory patterns can correlate to social cohesion, with, for example, less pro-social people more likely to flee away at times of conflict (Gilligan et al. 2014). Similarly, warfare tends to displace women relatively more than men,²⁷ ultimately altering social cohesion dynamics at the local level. In the present context, these sample selection issues are, however, unlikely to severely bias estimates. First, cell-time fixed effects should account for migration that is not ethnic-specific. Second, the episodes of ethnic violence observed in the sample stem mostly from low-intensity conflicts, which are unlikely to trigger mass-scale migration. In Appendix D I provide empirical support to these intuitions, showing that neither the population share nor the gender composition of ethnolinguistic groups, were significantly affected by warfare.

Another obvious threat to inference stems from reverse causality. While ethnic violence is a powerful element in the identity-building process, it might well be the case that more cohesive and fiercer cultural groups engage more often in conflict, having more powerful grievances and mobilisation structures. The ethnic \times period fixed effects partly dispel these concerns. The model compares people with the same ethnic background and during the same period: ethnic-broad self-selection into violence is effectively taken into account.

Despite the rich set of fixed effects, equation (1) can not provide es-

25. For example, Moscona and co-authors (2017 and 2020) find that ethnic groups with segmentary lineage social structures have lower levels of generalised trust and are more prone to conflict in sub-Saharan Africa. Exclusion from power, separation across national borders and traditional reputation mechanisms have also been pointed out as predictors of violence (Cederman et al. 2010, Michalopoulos and Papaioannou 2016, Cao et al. 2021, Fouka and Schlöpfer 2022).

26. The statistics come from the World Bank: <https://data.worldbank.org/indicator/VC.IDP.TOCV?locations=ZG>.

27. Globally, by the end of 2018, women and girls represented slightly more than half of the 41 million people living in a situation of internal displacement as a result of conflict and violence (Internal Displacement Monitoring Center 2020, p. 5).

timates that are causal in the sense of exploiting (conditional) random variation. Unobserved local ethnic-specific dynamics – such as access to power or temporal resource windfalls – may still bias inference by acting on the group social cohesion and its conflict propensity. A series of imperfect measures is adopted to lower these concerns. First, additional socio-economic and political variables are included in equation (1), thus partially controlling for the above mentioned confounding factors. Second, various subsamples are considered so as to make observations more comparable and to lower sample selection concerns. Most notably, as shown in Appendix D.1, all non-violent ethnolinguistic groups are excluded from the analysis, thus ruling out that results are driven by some omitted characteristic radically differing between peaceful and conflict-prone ethnolinguistic groups. Finally, but perhaps most importantly, results still hold when we consider far-away distant conflict, which is definitely less likely to be influenced by local political or economic shocks. As shown in Section 5, ethnic violence has a significant impact on ethnic identification and social cohesion even when exposure is only indirect. The exercise effectively lessens sample selection issues (respondents are not directly involved in remote conflict) and the confounding effect of local unobservables, hence strengthening a causal interpretation of the main results.

C Baseline results

Table 2 reports the estimation of the model over social identity and cohesion. The first regression includes only the three sets of geographic, temporal, and ethnic fixed effects not interacted between themselves. The estimated marginal impact of conflict on ethnic identification is positive. Yet, it has a small size and is statistically indistinguishable from zero. Including cell-period dummies in the second column increases the size of the coefficient by more than 50%, suggesting the existence of some confounding factor operating at the spatio-temporal level. Local economic development trends seem a likely culprit: while decreasing the chances of conflict, they can simultaneously positively correlate with both social cohesion and group identification. The third column adds ethnic-period fixed effects. The estimated coefficient almost doubles in size to 0.113. The increase of magnitude can again be explained in terms of omitted variable (*e.g.* access to political power) and underlines the importance of comparing observations belonging to the same cell \times period and ethnic \times period groups. The inclusion of Afrobarometer controls in column four only slightly decreases the magnitude of the co-

efficient: ethnic violence does indeed increase ethnic identification.

Table 2: OLS regressions of Ethnic identity & Social cohesion on Ethnic conflict

	(1)	(2)	(3)	(4)
ETHNIC IDENTITY				
Ethnic conflict	0.041 (0.037)	0.068 (0.034)**	0.113 (0.037)***	0.108 (0.037)***
Urban				-0.057 (0.014)***
R^2	0.13	0.21	0.22	0.22
N	103,646	103,605	103,588	101,378
SOCIAL COHESION				
Ethnic conflict	0.032 (0.025)	0.072 (0.026)***	0.076 (0.026)***	0.069 (0.026)***
Urban				-0.141 (0.009)***
R^2	0.16	0.22	0.22	0.27
N	109,417	109,388	109,371	107,094
Period FE	Yes	No	No	No
Cell FE	Yes	No	No	No
Ethnic FE	Yes	Yes	No	No
Cell-period FE.	No	Yes	Yes	Yes
Ethnic-period FE.	No	No	Yes	Yes
Individual cov.	No	No	No	Yes

Notes: The sample includes Afrobarometer rounds 3 to 6. Individual controls include: age and its square, gender, and an urban-rural indicator (*Urban*). The *Ethnic conflict* dummy signals the presence of conflict events involving the ethnolinguistic group of an observation and taking place in her cell of residence during the relevant period. *Ethnic Identity* captures the saliency of ethnic identification, while *Social Cohesion* is an index combining three dimensions of pro-sociality: attendance at community meetings, participation in instances of collective action, and membership in community associations. Standard errors are clustered at the cell-ethnic-period level and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

The estimated marginal impacts are both statistically and economically significant. In the most demanding specification, the coefficient is equal to 0.108, which is more than the half of the average difference between conflict-torn Uganda and the mostly-peaceful Zambian nation (0.199). To better gauge the relevance of conflict, compare it to the coefficient of the rural-urban indicator. Moving from the countryside to an urban centre decreases ethnic identification by 0.057, which is roughly

half of the effect triggered by ethnic violence.

The second panel of Table 2 reports results from equation (1) as estimated over the social cohesion index. The control of cell-period fixed effects seems particularly relevant in this context. Indeed, their inclusion substantially increases the magnitude of the coefficient, signalling the presence of some unobservable variable positively affecting community engagement and negatively correlating with conflict (*e.g.* local development trends). The point estimates of the last three columns are positive and statistically significant: ethnic conflict induces pro-social behaviour towards one's own community. Marginal effects are relatively modest when compared to the urban indicator, but comparable to the impact of conflict on ethnic identification: for both outcomes of interest, conflict exposure trigger an increase of about 9% standard deviations of the dependent variable.

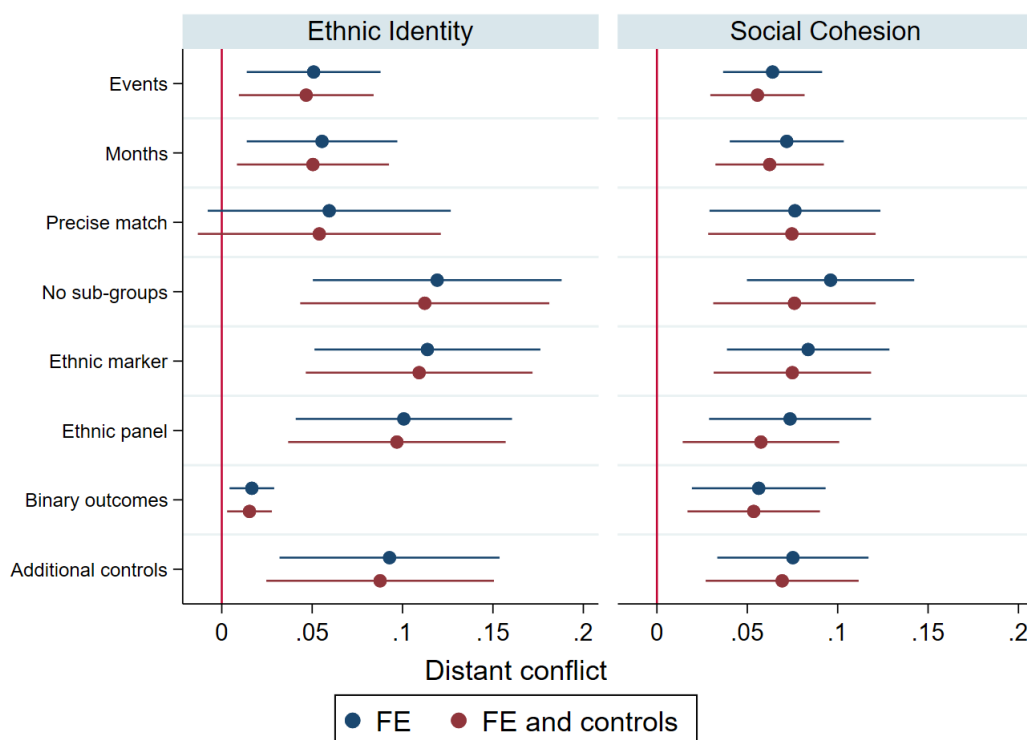
D Robustness checks

The models at equation (1) are perturbed over several dimensions to test their robustness.

First, sensitivity to alternative conflict definitions is explored. Two extensive measures of ethnic violence are used: the month- and event-based definitions introduced in Section 3. Having a lot of zeros, the two variables are parametrised using the inverse hyperbolic sine function; results are identical when natural logs are instead employed. Figure 3 shows their point estimates alongside their 90% confidence interval from regressions employing cell-period and ethnic-period fixed effects as well as the set of controls including the fixed personal characteristics of respondents. Estimates are positive and significant, signalling that ethnic identity and community engagement are responsive also to the intensity of violence.

A second battery of robustness tests concerns the match between ethnolinguistic groups and conflict actors. The two major issues encountered during the matching procedure were that: *(i)* the ethnic affiliation of the actor was not always blatantly clear, primarily because of its mixed objectives and identities; *(ii)* the associated ethnolinguistic group was a sub-group of the reported Afrobarometer language. Figure 3 reports point estimates and 90% confidence intervals from a version of equation (1) excluding these types of controversial matches. Two regressions for each match-outcome pair are run, one including only fixed effects, the other adding individual covariates. Results are overall robust to alternative matching procedures, the only exception being the marginal impact

Figure 3: OLS regressions of Ethnic identity & Social cohesion on Ethnic conflict – Robustness tests



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic conflict. Fixed effects correspond to cell \times period and ethnic \times period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. Standard errors are clustered at the cell-ethnic-period level. In the last panel (*Additional controls*), regressions labelled *FE* comprehend fixed effects, individual and socio-economic controls; regressions labelled *FE and controls* further add political covariates.

on ethnic identity that is no longer statistically different than zero when more imprecise matches are excluded.

A third series of robustness tests concerns the definition of the sample. First, ethnolinguistic groups not appearing in all four periods are dropped from the analysis. Indeed, their (dis-)appearance from the sample may be related to changes in their socio-political relevance. For example, Afrobarometer survey designers might have chosen to suddenly include a given ethnolinguistic group in light of its recent autonomy struggle. By excluding transient cultural entities, we avoid this type of selection issues. The second sample-wide modification concerns the dimension used for defining ethnolinguistic groups. The Afrobarometer question on ethnicity, rather than language, is thus employed.²⁸ As

28. On the reasons why the linguistic marker is preferable to ethnic's, see footnote 15.

shown in Figure 3, sign and significance of the baseline models are reproduced under these alternative samples.

The last robustness exercises concern the coding of outcome variables and the inclusion of additional covariates. In the literature, social cohesion measures are often transformed in binary indicators so as to avoid problems related to their possible subjective interpretation. I thus create dummy variables for ethnic identity and the three measures of pro-social behaviour, whose standardised score is then averaged into an alternative social cohesion index. The seventh panel of Figure 3 repeats this exercise: all point estimates are statistically significant and with same order of magnitude relatively to baseline estimates (7% standard deviations). I then include further covariates taken from Afrobarometer surveys. These variables control for the socio-economic profile of respondents as well as their political attitudes. In the first group are included: education, employment status, and an index of household wealth. Political controls comprehend: trust towards the president, trust towards the local government, and an indicator function for having voted in the last national election. Their inclusion aims at lowering the risks stemming from omitted variable issues; yet, they themselves might be endogenous in equation (1), and these results shall be read bearing this in mind. The last two regressions of Figure 3 are specifications of equation (1) cumulatively including socio-economic and political controls, beyond the usual set of fixed effects and individual-level covariates. All coefficients are positive, significant and comparable in size to the baseline estimates. Robustness to the inclusion of political controls is particularly relevant for the social identity outcome. Indeed, these covariates ensure that the observed increase in ethnic identification is not given by lower attachment to the nation-state, a process possibly triggered by higher conflict levels.

Appendix D shows robustness of the main results to additional sensitivity exercises, including: placebo analysis using future conflict activity, exclusion of the most peaceful and the most violent ethnolinguistic groups, sequential exclusion of each country.

5 Mechanisms

The empirical analysis in the previous section substantiated a significant and robust link between ethnic conflict and social cohesion. People exposed to ethnic violence reports higher attachment to their cultural group and exhibit greater pro-social behaviour, primarily aimed at their

community fellows. As seen in Section 1, the perhaps surprising result that conflict can prompt cooperative behaviour has been variously rationalised in the academic world. This section provides empirical support to the idea that conflict-induced ethnocentric dynamics are central for explaining the observed increase in social cohesion. Indeed, as one's own ethnolinguistic group is involved in some violent clash, the individual frames herself in ethnic terms, explicitly perceives her group identity, and rallies around it. This rally-around-the-flag effect is not limited to ethnic conflict and is witnessed across a wide range of settings where a group is faced by some external threat. Yet, what is remarkable about ethnic violence is the intensity of such phenomenon, capable of bringing about changes in actual behaviour even in the absence of instrumental incentives.²⁹ How can we make sense of this powerful shift in social preferences?

Various theories have been advanced to explain ethnocentrism, that is, the instinctive classification of the world into the superior ethnic 'us' and all the 'others' (Dennen 2000). Many of these theories have an explicit evolutionary framework in order to account for the massive evidence on human propensity to favourably treat co-ethnics (Haushofer et al. 2022). One of the most widely accepted theory understands ethnocentrism as residual, almost a manipulation, of natural selection on pro-kinship behaviour (Gat 2006).³⁰

Humans, as much as other animals, have a natural inclination to help their relatives. The behaviour is evolutionary sensible, inasmuch as family members are repositories of one's own genes and altruistic behaviour towards them does, indeed, ensure genetic reproduction. In the late Pleistocene – a time when human societies were relatively small and composed mainly of kin mates (Kelly 2007) – linguistic developments gradually led to ever-deepening cultural differentiation (Harari 2014). In this setting, common language and culture were effectively good indicators of kinship: favouring one's own co-ethnic was virtually equivalent to helping one's own relative. This is the socio-biological essence of ethnocentrism. In the pre-agricultural world, where ethnicity predicted

29. The peculiarities of ethnic warfare are more evident when other conflict types are considered. For example, Peri et al. (2021) empirically show that terrorist attacks in Europe have not a robust effect on trust towards government institutions and on other political attitudes. Interestingly, the authors find that only *foreign* terrorist attacks have the effect of increasing the appeal of *nationalistic* parties.

30. A thorough treatment of ethnocentrism, discussing also alternative theories such as those advanced by Bowles (2008, 2009, 2012) and co-authors (Choi and Bowles 2007), is relegated to Appendix C.

kinship, it is easy to see the evolutionary rationale of the us-them discrimination. Yet, as human societies grew larger and more complex, the application of kin-solidarity to culturally related people (*i.e.* co-ethnics) definitely lost any evolutionary aim. When during the Second World War Japanese kamikazes committed suicide against US ships, they were neither increasing their biological fitness, nor that of any distant relative. Nonetheless, some of them were volunteers, enthusiastically going to risk their lives for the *motherland*. This seemingly irrational behaviour shall be read as a manipulation, an extension beyond its original objective, of kin-based solidarity within small groups.³¹ It is this the atavistic component of contemporary xenophobia, the same often perplexing observers of ethnic cleansing and genocides.

Ethnic conflict thus plays onto human ethnocentric instincts, activating feelings of solidarity and altruism even in the absence of tangible economic benefits.

A Identifying the ethnocentric channel

The use of ethnic-varying conflict and the positive link found between this and ethnic identification, already suggest that the ethnocentric channel is central for explaining the observed increase in pro-social behaviour. The remainder of the paragraph carries out additional tests of this mechanism.

First, if a process of ethnic radicalisation drives the results, we would expect that people are responsive also to distant conflict involving their co-ethnics. Table 3 reports estimates of a model employing conflict spatially aggregated over the eight-cell neighbourhood of each respondent. We are thus considering ethnic violence outside of the cell of the respondent, taking place up to 155km away from each observation. The first two columns include only this external measure of conflict, while the last two add also the baseline local conflict indicator, which otherwise is a plausible omitted variable. The point estimate of distant conflict is always positive and mostly statistically significant, albeit having a magnitude lower than that of local violence. Interestingly, as shown in Appendix E, ethnic conflict has still a positive impact on social identity and cohesion, when it is spatially aggregated over the 24-cell neighbourhood of respondents.

31. The persistence of naturally selected behaviours even if devoid of their evolutionary *raison d'être*, albeit surprising, is typical of humankind. For example, in the same logic, as humans we have a natural love for sweet food, even though sweetness does not serve any more its original evolutionary aim of signalling fruit maturity.

Table 3: OLS regressions of Ethnic identity & Social cohesion on Distant ethnic conflict

	(1)	(2)	(3)	(4)
ETHNIC IDENTITY				
Distant conflict	0.058 (0.030)*	0.057 (0.030)*	0.049 (0.030)*	0.049 (0.030)
Ethnic conflict			0.107 (0.037)***	0.101 (0.037)***
R^2	0.22	0.22	0.22	0.22
N	103,588	101,378	103,588	101,378
SOCIAL COHESION				
Distant conflict	0.059 (0.021)***	0.044 (0.020)**	0.053 (0.021)**	0.039 (0.020)*
Ethnic conflict			0.069 (0.027)***	0.063 (0.026)**
R^2	0.22	0.27	0.22	0.27
N	109,371	107,094	109,371	107,094
Fixed effects	Yes	Yes	Yes	Yes
Individual cov.	No	Yes	No	Yes

Notes: The sample includes Afrobarometer rounds 3 to 6. Fixed effects correspond to cell \times period and ethnic \times period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. The *Ethnic conflict* dummy signals the presence of conflict events involving the ethnolinguistic group of an observation and taking place in her cell of residence during the relevant period. *Distant conflict* is a dummy variable signalling the presence of ethnic conflict in the 8-cell neighbourhood of an observation. *Ethnic Identity* captures the saliency of ethnic identification, while *Social Cohesion* is an index combining three dimensions of pro-sociality: attendance at community meetings, participation in instances of collective action, and membership in community associations. Standard errors are clustered at the cell-ethnic-period level and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

The importance of this exercise is twofold. On the one hand, it lowers the endogeneity issues mentioned in Section 4B: local ethnic-specific shocks are unlikely to be correlated to distant clashes. On the other hand, it pins down ethnocentrism as the causal force behind the results of the main empirical analysis. Indeed, alternative mechanisms are ill-suited to explain how remote ethnic violence can trigger higher social cohesion. For example, conflict-induced economic incentives to cooperation work primarily on a geographical base: it is hard to see how local insurance schemes are adopted in response to rumours of distant conflict. Other psychological mechanisms, such as post-traumatic growth,

are also usually understood as taking place in the midst of a direct violent trauma.

A second strategy to test the relevance of the ethnocentric channel is to compare ethnic and non-ethnic conflict, whereby only the first should have a significant impact on social identification and cohesion if identity-based dynamics are at play. Define $Z_{-e,k,t}$ as a dummy taking value of one for conflict activity taking place in cell k at period t and unrelated to ethnolinguistic group e . Figure 4 reports point estimates of $Z_{-e,k,t}$ and $C_{e,k,t}$, with their 90% confidence intervals, as retrieved from the following model:

$$y_{i,e,k,t} = \mu_{k,t} + \lambda_{e,t} + \beta C_{e,k,t} + \psi Z_{-e,k,t} + \delta' x_{i,e,k,t} + u_{i,e,k,t} \quad (2)$$

Note that baseline regressions are already estimating the effect of ethnic conflict net of the total warfare present in each cell-period, as captured by the spatio-temporal fixed effects $\mu_{k,t}$. Regression (2) contrast ethnic to non-ethnic violence by comparing cases where only one of the two is present to cases where both of them are taking place. The 'horse-race' between ethnic and non-ethnic conflict is clearly won by the first measure of warfare: the coefficient of $Z_{-e,k,t}$ is always negative and insignificantly different than zero. The point estimate of β , instead, remains positive and significant.

Finally, a more direct test of the ethnocentric channel can be easily carried out using an index of ethnolinguistic polarisation. This is defined as:

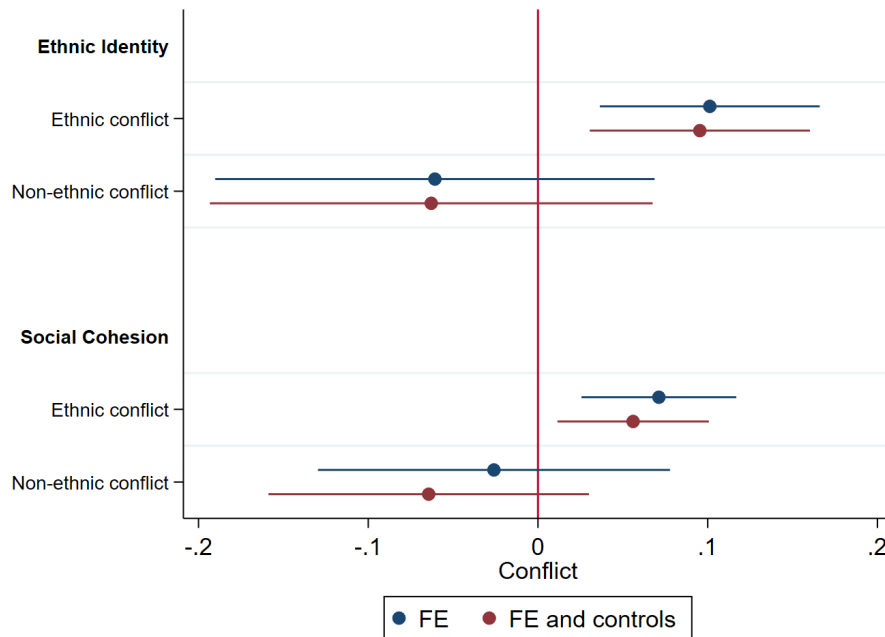
$$ELP_{k,t} = 1 - \sum_e \left(\frac{\frac{1}{2} - \pi_{e,k,t}}{\frac{1}{2}} \right)^2 \pi_{e,k,t}$$

Where $\pi_{e,k,t}$ is the population share of ethnic group e in cell k at time t . The index measures how far the distribution of ethnic groups is from the bipolar distribution. As $ELP_{k,t}$ moves to one, polarisation increases, signalling that the cell-period is evenly divided between two groups. In areas marked by a high ethnolinguistic polarisation, local public goods are contested between equally powerful entities and cooperation at the community level goes benefiting people hailing from different cultural backgrounds. Hence, in these zones, if conflict affects social cohesion through ethnocentric dynamics, the estimated positive impact should be lower.

We thus define a new model by augmenting equation (1) with the ethnolinguistic polarisation index and its interaction with conflict. Formally:

$$y_{i,e,k,t} = \mu_{k,t} + \lambda_{e,t} + \beta C_{e,k,t} + \alpha(C_{e,k,t} \times ELP_{k,t}) + \delta' x_{i,e,k,t} + u_{i,e,k,t} \quad (3)$$

Figure 4: OLS regressions of Social identity and cohesion on Ethnic and non-ethnic conflict



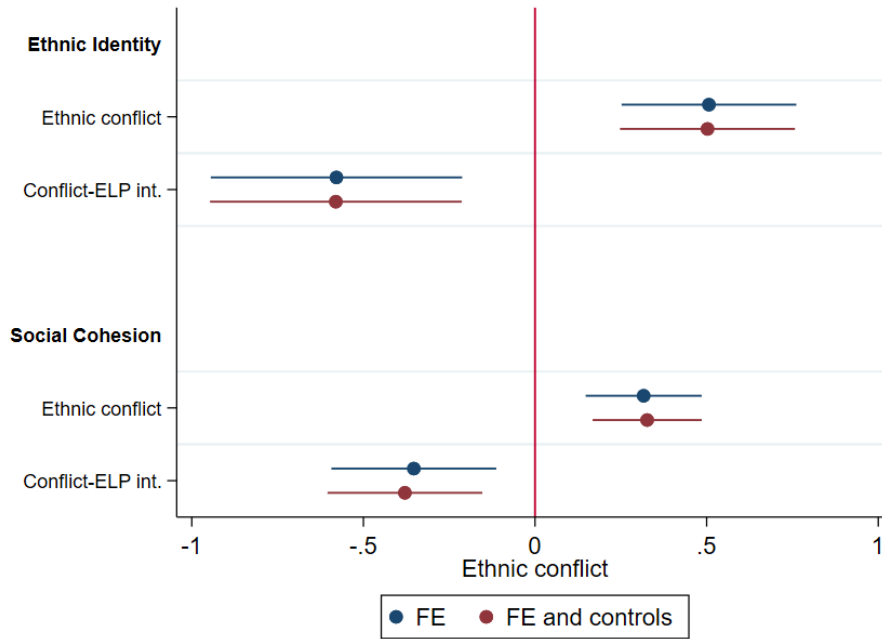
Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic and non-ethnic conflict. Fixed effects correspond to cell \times period and ethnic \times period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. Standard errors are clustered at the cell-ethnic-period level.

Note that the term $ELP_{k,t}$ is absorbed by the $\mu_{k,t}$ fixed effects. If conflict-induced pro-social behaviour remains limited to one's own co-ethnics, such a cooperation should be harder to achieve in districts more polarised. That is, the joint impact of ethnic polarisation and warfare (α) shall be negative.

Figure 5 reports estimates of equation (3). While conflict still has a positive impact on community engagement and ethnic identification, its interaction with the ELP is always significantly negative. There are, indeed, heterogeneous effects with respect to the ethnolinguistic composition of the various cell-period observations. In areas with large levels of ethnolinguistic polarisation, the positive impact of ethnic conflict decreases for the very simple fact that within-community cooperation goes benefiting also people from other cultural units. Put it differently, the pro-social impact of conflict is *parochial*, being directed only towards one's own co-ethnics.

As shown in Appendix E, the results of this section are robust to: employing extensive measures of ethnic conflict, different actor-ethnicity

Figure 5: Heterogeneous effects by Ethnolinguistic Polarisation



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic conflict and its interaction with an index of ethnolinguistic polarisation. Fixed effects correspond to cell×period and ethnic×period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. Standard errors are clustered at the cell-ethnic-period level.

matching rules, the use of alternative sample restrictions, a binary re-coding of outcome variables, and the inclusion of socio-economic and political covariates.

6 Conclusions

This paper has enquired the impact of ethnic conflict on social identity and norms. The setting of the empirical analysis is sub-Saharan Africa, where sub-national identities are still very alive and affect the course of public policy. The research design relies on a novel manually-compiled dataset, linking conflict actors to ethnolinguistic groups. The procedure allowed the construction of conflict measures varying at the spatial, temporal, and ethnic level. This triple source of variation and the ensuing rich set of fixed effects, are at the core of the identification strategy here employed.

Confirming the anecdotal evidence on conflict-induced identity radicalisation, this paper substantiates the first cross-country empirical link

between ethnic conflict and ethnic identification. This result is then used to shed lights on the second empirical contribution of this article: ethnic violence prompts cooperative behaviour in communities exposed to it. The positive conflict-identity relationship found and the explicit use of ethnic, rather than geographical, warfare suggest that identity-based psychological processes are at the base of the observed increase in social cohesion. This intuition is further developed and tested. First, *distant* ethnic violence positively affects civic attitudes, making improbable that results are driven by direct war traumas or by the effect of conflict on local institutions. Second, conflict unrelated to the ethnolinguistic group of respondents has not any impact on their community engagement, making clear that the ethnic dimension of conflict is what really matters. Finally, conflict-induced pro-social behaviour is conditional on low levels of ethnolinguistic polarisation, indicating that cooperation likely remains limited to one's own group.

A socio-biological understanding of ethnocentrism provides a good theoretical framework to make sense of these findings. Conflict along ethnic cleavages plays, indeed, onto the innate 'us-them' ethnic division of the world, which can neither be dismissed as fully artificial nor as completely genetic. It is, indeed, in light of this complex dual nature that both out-group xenophobia and in-group cooperation shall be read. This paper found evidence on the second, giving empirical support to the idea that, regardless of instrumental considerations, feelings of identification and solidarity towards one's own ethnic 'us' can lead to actual pro-social behaviour, albeit in a parochial fashion.

References

- Acemoglu, Daron, and James A Robinson. 2012. *Why nations fail: The origins of power, prosperity and poverty*. Profile London.
- Allen, Tim, and Jean Seaton. 1999. *The media of conflict: War reporting and representations of ethnic violence*. Zed Books.
- Ananyev, Maxim, and Mikhail Poyker. 2021. "State Capacity and Demand for Identity: Evidence from Political Instability in Mali." *Working paper*, ISSN: 1556-5068. <https://doi.org/10.2139/ssrn.3419294>.
- Asuni, Judith Burdin. 2009. *Understanding the Armed Groups of the Niger Delta*. Working Paper. Council on Foreign Relations.
- Barbero, Alessandro. 2009. *I conflitti d'identità nella storia*. Online available at: https://www.youtube.com/watch?v=h6C_qI347Bo&ab_channel=FestivaLEconomia [Accessed 30/01/2022].
- Bateson, Regina. 2012. "Crime victimization and political participation." *American Political Science Review* 106 (3): 570–587.
- Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts. 2016. "Can war foster cooperation?" *Journal of Economic Perspectives* 30 (3): 249–74.
- Bellows, John, and Edward Miguel. 2006. "War and institutions: New evidence from Sierra Leone." *American Economic Review* 96 (2): 394–399.
- . 2009. "War and local collective action in Sierra Leone." *Journal of public Economics* 93 (11-12): 1144–1157.
- Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig. 2017. "This mine is mine! How minerals fuel conflicts in Africa." *American Economic Review* 107 (6): 1564–1610.
- Bernhard, Helen, Urs Fischbacher, and Ernst Fehr. 2006. "Parochial altruism in humans." *Nature* 442 (7105): 912–915.
- Besley, Timothy, and Marta Reynal-Querol. 2014. "The legacy of historical conflict: Evidence from Africa." *American Political Science Review* 108 (2): 319–336.
- Blattman, Christopher. 2009. "From violence to voting: War and political participation in Uganda." *American political Science review* 103 (2): 231–247.
- Boone, Catherine, and Lydia Nyeme. 2015. "Land institutions and political ethnicity in Africa: evidence from Tanzania." *Comparative Politics* 48 (1): 67–86.

- Bowles, Samuel. 2008. "Being human: Conflict: Altruism's midwife." *Nature* 456 (7220): 326–327.
- . 2009. "Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors?" *Science* 324 (5932): 1293–1298.
- . 2012. "Warriors, levelers, and the role of conflict in human social evolution." *Science* 336 (6083): 876–879.
- Brewer, Marilynn B. 2007. "The importance of being we: Human nature and intergroup relations." *American psychologist* 62 (8): 728.
- Bringa, Tone, and Debbie Christie. 1993. *We are all neighbours*. Disappearing World Series.
- Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics* 225 (2): 200–230.
- Cao, Yiming, Benjamin Enke, Armin Falk, Paola Giuliano, and Nathan Nunn. 2021. *Herding, Warfare, and a Culture of Honor: Global Evidence*. NBER Working Papers. National Bureau of Economic Research.
- Cassar, Alessandra, Pauline Grosjean, and Sam Whitt. 2013. "Legacies of violence: trust and market development." *Journal of Economic Growth* 18 (3): 285–318.
- Cederman, Lars-Erik, Andreas Wimmer, and Brian Min. 2010. "Why do ethnic groups rebel? New data and analysis." *World politics* 62 (1): 87–119.
- Choi, Jung-Kyoo, and Samuel Bowles. 2007. "The coevolution of parochial altruism and war." *science* 318 (5850): 636–640.
- De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110 (9): 2964–96.
- De Juan, Alexander, and Jan Henryk Pierskalla. 2016. "Civil war violence and political trust: Microlevel evidence from Nepal." *Conflict Management and Peace Science* 33 (1): 67–88.
- De Luca, Giacomo, and Marijke Verpoorten. 2015a. "Civil war and political participation: Evidence from Uganda." *Economic Development and Cultural Change* 64 (1): 113–141.
- . 2015b. "Civil war, social capital and resilience in Uganda." *Oxford Economic Papers* 67 (3): 661–686.

- Dennen, Johan M G van der. 2000. "Of badges, bonds and boundaries: In-group/outgroup differentiation and ethnocentrism revisited." In *In-group/outgroup behaviour in modern societies. An evolutionary perspective*, edited by K. Thienpont and R. Cliquet, 37–74. NJCM Boekerij.
- Depetris-Chauvin, Emilio, Ruben Durante, and Filipe Campante. 2020. "Building nations through shared experiences: Evidence from African football." *American Economic Review* 110 (5): 1572–1602.
- Easterly, William, and Ross Levine. 1997. "Africa's growth tragedy: policies and ethnic divisions." *The quarterly journal of economics*, 1203–1250.
- Eifert, Benn, Edward Miguel, and Daniel N Posner. 2010. "Political competition and ethnic identification in Africa." *American journal of political science* 54 (2): 494–510.
- Elischer, Sebastian. 2013. *Political parties in Africa: Ethnicity and party formation*. Cambridge University Press.
- Fearon, James D, and David D Laitin. 2003. "Ethnicity, insurgency, and civil war." *American political science review* 97 (1): 75–90.
- Fouka, Vasiliki, and Alain Schlöpfer. 2022. *Culture Clash: Incompatible Reputation Mechanisms and Intergroup Conflict*. Technical report. National Bureau of Economic Research.
- Gat, Azar. 2006. *War in human civilization*. Oxford University Press.
- Gilligan, Michael J, Benjamin J Pasquale, and Cyrus Samii. 2014. "Civil war and social cohesion: Lab-in-the-field evidence from Nepal." *American Journal of Political Science* 58 (3): 604–619.
- Grosjean, Pauline. 2014. "Conflict and social and political preferences: Evidence from World War II and civil conflict in 35 European countries." *Comparative Economic Studies* 56 (3): 424–451.
- Guarnieri, Eleonora, and Ana Tur-Prats. 2022. "Cultural distance and conflict-related sexual violence." *Working paper*.
- Guriev, Sergei, and Nikita Melnikov. 2016. "War, inflation, and social capital." *American Economic Review* 106 (5): 230–35.
- Hamilton, Annette. 2006. "Reflections on the 'disappearing Sakai': a tribal minority in southern Thailand." *Journal of Southeast Asian Studies* 37 (2): 293–314.
- Harari, Yuval Noah. 2014. *Sapiens: A brief history of humankind*. Random House.

- Hartman, Alexandra C, and Benjamin S Morse. 2020. "Violence, empathy and altruism: evidence from the Ivorian refugee crisis in Liberia." *British Journal of Political Science* 50 (2): 731–755.
- Haushofer, Johannes, Sara Lowes, Abednego Musau, David Ndeti, Nathan Nunn, Moritz Poll, and Nancy Qian. 2022. *Stress, Ethnicity, and Prosocial Behavior*. NBER Working Papers. National Bureau of Economic Research.
- Hazen, Jennifer M, and Jonas Horner. 2007. *Small arms, armed violence, and insecurity in Nigeria: The Niger Delta in perspective*. JSTOR.
- Hendrix, Cullen S, and Idean Salehyan. 2012. "Climate change, rainfall, and social conflict in Africa." *Journal of peace research* 49 (1): 35–50.
- Herbst, Jeffrey. 2000. *States and power in Africa*. Princeton University Press.
- Hopkins, Antony G. 2019. *An economic history of West Africa*. Routledge.
- Internal Displacement Monitoring Center. 2020. *Women and Girls in Internal Displacement*. Online available at: <https://www.internal-displacement.org/sites/default/files/publications/documents/202003-twice-invisible-internally-displaced-women.pdf> [Accessed 30/01/2022].
- Jennings, Colin, and Santiago Sanchez-Pages. 2017. "Social capital, conflict and welfare." *Journal of Development Economics* 124:157–167.
- Kalyvas, Stathis N. 2008. "Ethnic defection in civil war." *Comparative Political Studies* 41 (8): 1043–1068.
- Kaufmann, Chaim. 1996. "Possible and impossible solutions to ethnic civil wars." *International security* 20 (4): 136–175.
- Kelly, Robert L. 2007. *The foraging spectrum: Diversity in hunter-gatherer lifestyles*. Smithsonian Inst Press.
- Kesselman, Mark, Joel Krieger, and William A Joseph. 2016. *Introduction to comparative politics: political challenges and changing agendas*. 7th ed. Cengage Learning.
- Kibris, Arzu, and Resul Cesur. 2022. *War Does not Foster Cooperation when Neoclassical Mechanisms are Absent! Evidence from a Natural Experiment among Turkish Conscripts*. NBER Working Papers. National Bureau of Economic Research.
- Laitin, David D. 2004. "Ethnic unmixing and civil war." *Security Studies* 13 (4): 350–365.
- Lee, Richard Borshay. 1979. *The !Kung San: men, women and work in a foraging society*. Cambridge University Press.

- Lomo, Zachary, and Lucy Hovil. 2004. "Behind the violence. The war in northern Uganda." *Institute for Security Studies Monographs* 2004 (99): 87.
- Manacorda, Marco, and Andrea Tesei. 2020. "Liberation technology: Mobile phones and political mobilization in Africa." *Econometrica* 88 (2): 533–567.
- Michalopoulos, Stelios, and Elias Papaioannou. 2013. "Pre-colonial ethnic institutions and contemporary African development." *Econometrica* 81 (1): 113–152.
- . 2016. "The long-run effects of the scramble for Africa." *American Economic Review* 106 (7): 1802–48.
- Miguel, Edward. 2004. "Tribe or nation? Nation building and public goods in Kenya versus Tanzania." *World politics* 56 (3): 327–362.
- Mironova, Vera, and Sam Whitt. 2016. "The evolution of prosociality and parochialism after violence." *Journal of Peace Research* 53 (5): 648–664.
- . 2018. "Social norms after conflict exposure and victimization by violence: Experimental evidence from Kosovo." *British Journal of Political Science* 48 (3): 749–765.
- . 2021. "Conflict and parochialism among combatants and civilians: Evidence from Ukraine." *Journal of Economic Psychology* 86:102425.
- Moscona, Jacob, Nathan Nunn, and James A Robinson. 2017. "Keeping it in the family: Lineage organization and the scope of trust in sub-Saharan Africa." *American Economic Review* 107 (5): 565–71.
- . 2020. "Segmentary Lineage Organization and Conflict in Sub-Saharan Africa." *Econometrica* 88 (5): 1999–2036.
- Mueller, John. 2000. "The banality of "ethnic war"." *International Security* 25 (1): 42–70.
- Müller-Crepon, Carl, Yannick I Pengl, and Nils-Christian Bormann. 2020. "Linking ethnic data from africa." *Unpublished Manuscript. ETH Zurich*. https://yannickpengl.com/src/research/ethnic_matching.pdf.
- Nunn, Nathan, and Leonard Wantchekon. 2011. "The slave trade and the origins of mistrust in Africa." *American Economic Review* 101 (7): 3221–52.
- Olson, James Stuart. 1996. *The peoples of Africa: an ethnohistorical dictionary*. Greenwood Publishing Group.
- Oriola, Temitope B. 2013. *Criminal Resistance?: The Politics of Kidnapping of Oil Workers (Interdisciplinary Research Series in Ethnic, Gender and Class Relations)*. Ashgate Publishing Group.

- Peri, Giovanni, Daniel I Rees, and Brock Smith. 2021. *Terrorism and Political Attitudes: Evidence from European Social Surveys*. NBER Working Papers 28662. National Bureau of Economic Research.
- Peterson, Scott. 2000. *Me against my brother: at war in Somalia, Sudan and Rwanda*. Routledge.
- Pinker, Steven. 2012. *The false allure of group selection*. Online available at: <https://www.edge.org/conversation/stevenpinker-the-false-allure-of-group-selection> [Accessed 30/01/2022].
- Pohl, Walter. 2015. "Political uses of ethnicity in early medieval Europe." In *Ethnicity as a Political Resource: Conceptualizations across Disciplines, Regions, and Periods*, edited by University of Cologne Forum, 201–208. Transcript Verlag.
- Posner, Daniel N. 2004. "The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi." *American Political Science Review* 98 (4): 529–545.
- Powell, Steve, Rita Rosner, Willi Butollo, Richard G Tedeschi, and Lawrence G Calhoun. 2003. "Posttraumatic growth after war: A study with former refugees and displaced people in Sarajevo." *Journal of clinical psychology* 59 (1): 71–83.
- Resnik, Julia. 2003. "'Sites of memory' of the Holocaust: shaping national memory in the education system in Israel." *Nations and nationalism* 9 (2): 297–317.
- Robinson, Amanda Lea. 2014. "National versus ethnic identification in Africa: Modernization, colonial legacy, and the origins of territorial nationalism." *World Politics* 66 (4): 709–746.
- Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti. 2013. "Seeds of distrust: Conflict in Uganda." *Journal of Economic Growth* 18 (3): 217–252.
- Sambanis, Nicholas, and Moses Shayo. 2013. "Social identification and ethnic conflict." *American Political Science Review* 107 (2): 294–325.
- Scott, James C. 2009. *The art of not being governed*. Yale University Press.
- . 2017. *Against the grain*. Yale University Press.
- Seter, Hanne. 2016. "Connecting climate variability and conflict: Implications for empirical testing." *Political Geography* 53:1–9.
- Shayo, Moses, and Asaf Zussman. 2017. "Conflict and the persistence of ethnic bias." *American Economic Journal: Applied Economics* 9 (4): 137–65.

- Simiyu, Robert Romborah. 2008. "Militianisation of resource conflicts: the case of land-based conflict in the Mount Elgon region of western Kenya." *Institute for Security Studies Monographs*, no. 152, 80.
- Tedeschi, Richard G, and Lawrence G Calhoun. 2004. "Posttraumatic growth: conceptual foundations and empirical evidence." *Psychological inquiry* 15 (1): 1–18.
- Tilly, Charles. 1994. "States and nationalism in Europe 1492-1992." *Theory and society*, 131–146.
- Voors, Maarten J, and Erwin H Bulte. 2014. "Conflict and the evolution of institutions: Unbundling institutions at the local level in Burundi." *Journal of Peace Research* 51 (4): 455–469.
- Welsh, David. 1996. "Ethnicity in sub-saharan Africa." *International Affairs* 72 (3): 477–491.
- Wrangham, Richard. 2019. *The goodness paradox: The strange relationship between virtue and violence in human evolution*. Vintage.
- Yamagishi, Toshio, and Nobuhiro Mifune. 2016. "Parochial altruism: Does it explain modern human group psychology?" *Current Opinion in Psychology* 7:39–43.

**Online Appendix to
Identity conflict, ethnocentrism and social
cohesion**

Matteo Sestito[†]

[†]. Aix-Marseille University, CNRS, EHESS, Centrale Marseille, AMSE, France.

Contents

A Data	1
A.1 <i>Outcome and control variables</i>	2
A.2 <i>Conflict variables</i>	5
B Africa ethnolinguistic armed groups Dataset	6
B.1 <i>Dataset coverage</i>	7
B.2 <i>Descriptive statistics</i>	9
C Ethnocentrism	11
D Baseline - Robustness tests	14
D.1 <i>Outlier observations</i>	14
D.2 <i>Standard error clustering</i>	18
D.3 <i>Alternative treatment and outcome specifications</i>	19
D.4 <i>Placebo</i>	21
D.5 <i>Migration</i>	22
E Mechanisms - Robustness tests	25
E.1 <i>Distant conflict</i>	25
E.2 <i>Ethnic vs non-ethnic conflict</i>	26
E.3 <i>Heterogeneous effects</i>	27

A Data

This section provides further details on the data employed in the empirical analysis of the paper. Table A1 displays summary statistics for all the variables used, while the subsections below report further information for outcome, control, and conflict variables.

Table A1: Descriptive statistics

	Mean	Sd	Min	Max	<i>N</i>
CONFLICT					
Ethnic conflict	0.043	0.203	0	1	111428
Distant conflict	0.096	0.295	0	1	111428
Remote conflict	0.135	0.342	0	1	111428
Non-ethnic conflict	0.340	0.474	0	1	111428
Conflict-ELP	0.022	0.119	0	1	111428
Events	0.134	1.161	0	33	111428
Months	0.045	0.223	0	1	111428
Precise match	0.034	0.181	0	1	111428
No sub-groups	0.035	0.183	0	1	111428
OUTCOMES					
Ethnic Identity	2.362	1.196	1	5	103656
Social Cohesion	0.002	0.791	-1.25	1.89	109426
Att. Comm. Meetings	2.150	1.302	0	4	110772
Collective Action	1.855	1.293	0	4	110520
Community Associations	0.609	0.928	0	3	110387
CONTROLS					
Age	36.810	14.772	18	130	110350
Male	0.498	0.500	0	1	111427
Urban residence	0.385	0.487	0	1	110158
Education	3.298	1.997	0	9	111173
Employment status	1.277	1.157	0	3	111019
Household wealth	6.039	4.617	0	20	110015
Voted	0.724	0.447	0	1	109664
Trust president	1.826	1.127	0	3	107208
Trust local govt.	1.566	1.074	0	3	102953

Notes: The *Events* and *Months* specifications of conflict are in levels.

A.1 Outcome and control variables

Individual-level information on social norms comes from Afrobarometer surveys.¹ Here is reported the exact text of Afrobarometer variables alongside the meaning of their values.²

- Ethnic identity
 - *Afrobarometer question*: “Let us suppose that you had to choose between being a [enter nationality] and being a [Respondent’s ethnic group]. Which of the following best expresses your feelings?”
 - *Values (recoded)*: 1=I feel only [enter nationality]; 2=I feel more [enter nationality] than (Respondent’s ethnic group); 3=I feel equally [enter nationality] and (Respondent’s ethnic group); 4=I feel more (Respondent’s ethnic group) than [enter nationality]; 5=I feel only (Respondent’s ethnic group)
 - *Values in binary recoding*: 0=Otherwise, 1=I feel only (Respondent’s ethnic group)
 - *Afrobarometer question number (rounds 3 to 6)*: Q82, Q83, Q85b, Q88b
 - *Further comments*: The original coding of the variable is inverted around the median value of 3 (e.g. value 1 in the original coding corresponds to value 5 in the recoded variable used in the analysis)
- Attendance community meetings
 - *Afrobarometer question*: “Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year. If not, would you do this if you had the chance: Attended a community meeting?”
 - *Values*: 0=No, would never do this; 1=No, but would do if had the chance; 2=Yes, once or twice; 3=Yes, several times; 4=Yes, often
 - *Values in binary recoding*: 0=Otherwise; 1=Yes, often
 - *Afrobarometer question number (rounds 3 to 6)*: Q31a, Q23a, Q26a, Q20a
- Collective action
 - *Afrobarometer question*: “Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year. If not,

1. The surveys are publicly available at: <http://afrobarometer.org/data/geocoded-data>.

2. The exact wording of the question is taken from the sixth Afrobarometer round.

- would you do this if you had the chance: Got together with others to raise an issue?”
- *Values:* 0=No, would never do this; 1=No, but would do if had the chance; 2=Yes, once or twice; 3=Yes, several times; 4=Yes, often
 - *Values in binary recoding:* 0=Otherwise; 1=Yes, often
 - *Afrobarometer question number (rounds 3 to 6):* Q31b, Q23b, Q26b, Q20b
- **Membership community associations**
 - *Afrobarometer question:* “Let’s turn to your role in the community. Now I am going to read out a list of groups that people join or attend. For each one, could you tell me whether you are an official leader, an active member, an inactive member, or not a member: Some other voluntary association or community group?”
 - *Values:* 0=Not a member; 1=Inactive member; 2=Active member; 3=Official leader
 - *Values in binary recoding:* 0=Not a member & Inactive member; 1=Otherwise
 - *Afrobarometer question number (rounds 3 to 6):* Q28d, Q22b, Q25b, Q19b
 - *Further comments:* The question in the third round is phrased slightly differently. Moreover, in this survey, contrarily to the others, respondents were previously asked also for their membership in trade unions and business associations
 - **Age**
 - *Afrobarometer question:* “How old are you?”
 - *Values:* 18-130
 - *Afrobarometer question number (rounds 3 to 6):* Q1, Q1, Q1, Q1
 - **Gender**
 - *Afrobarometer question:* “Respondent’s gender”
 - *Values (recoded):* 0= Female; 1= Male
 - *Afrobarometer question number (rounds 3 to 6):* Q101, Q101, Q101, Q101
 - **Urban**
 - *Afrobarometer question:* “Urban or Rural Primary Sampling Unit”
 - *Values (recoded):* 0=Rural; 1=Urban
 - *Afrobarometer question number (rounds 3 to 6):* Urbrur
 - **Education**

- *Afrobarometer question*: “What is your highest level of education?”
 - *Values*: 0=No formal schooling; 1=Informal schooling only (including Koranic schooling); 2=Some primary schooling, 3=Primary school completed; 4=Intermediate school or Some secondary school / high school; 5=Secondary school / high school completed, 6=Post-secondary qualifications, other than university *e.g.* a diploma or degree from a polytechnic or college; 7=Some university; 8=University completed; 9=Post-graduate
 - *Afrobarometer question number (rounds 3 to 6)*: Q90, Q89, Q97, Q97
- **Employment status**
 - *Afrobarometer question*: “Do you have a job that pays a cash income? If yes, is it full-time or part-time? If no, are you presently looking for a job?”
 - *Values (recoded)*: 0=No (not looking); 1=No (looking); 2=Yes part time; 3= Yes, full time
 - *Afrobarometer question number (rounds 3 to 6)*: Q94, Q94, Q96, Q95
 - *Further comments*: In Afrobarometer rounds 3 and 4 respondents can distinguish between “not looking” and “looking” also in reference to the part- and full-time categories. This distinction is eliminated to express the variables as in the value-scale reported above
- **Household wealth**
 - The covariate is equal to the average of five standardised variables measuring household wealth
 - *Afrobarometer question*: “Over the past year, how often, if ever, have you or anyone in your family gone without: *X*?”
Where $X \in \{\text{enough food to eat, enough clean water for home use, medicines or medical treatment, enough fuel to cook your food, a cash income}\}$
 - *Values*: 0=Never; 1=Just once or twice; 2=Several times; 3=Many times; 4=Always
 - *Afrobarometer question number (rounds 3 to 6)*: Q8a-e, Q8a-e, Q8a-e, Q8a-e
- **Trust president**
 - *Afrobarometer question*: “How much do you trust each of the following, or haven’t you heard enough about them to say: The President?”
 - *Values*: 0=Not at all; 1=Just a little; 2=Somewhat; 3=A lot
 - *Afrobarometer question number (rounds 3 to 6)*: Q55a, Q49a, Q59a, Q52a
- **Trust local government**

- *Afrobarometer question*: “How much do you trust each of the following, or haven’t you heard enough about them to say: Your Metropolitan, Municipal or District Assembly?”
 - *Values*: 0=Not at all; 1=Just a little; 2=Somewhat; 3=A lot
 - *Afrobarometer question number (rounds 3 to 6)*: Q55d, Q49d, Q59e, Q52e
- Voted
 - *Afrobarometer question*: “Understanding that some people were unable to vote in the most recent national election in [20xx], which of the following statements is true for you?”
 - *Values (recoded)*: 0=Did not vote; 1=Voted
 - *Afrobarometer question number (rounds 3 to 6)*: Q30, Q23d, Q27, Q21
 - *Further comments*: The original question list various reasons for not having voted, which change from one Afrobarometer round to another. They all have been coded within the same no-vote category

A.2 Conflict variables

Ethnic conflict variables are built starting from the ACLED dataset, which provides information on the location, date, and actors involved in conflict incidents spanning the whole African continent since 1997. I focus only on non-state armed groups. These are the actors defined by ACLED as rebel groups, political and identity militias: interaction codes 2, 3, and 4, respectively. I thus assign, where relevant, an ethnic affiliation to these armed groups (more details on the *Africa ethnolinguistic armed groups Dataset* in the next section). The procedure yields a cross-section of conflict incidents $c_{e,l,j}$ taking place in location l , at day j , and involving an armed group associated to ethnolinguistic group e . The baseline measure excludes conflict incidents where a state force is involved.

Then conflict incidents are aggregated over space and time. Spatially, I consider all the locations l belonging to a cell k . Temporally, I aggregate over the period between two consecutive Afrobarometer surveys, that is, over the day-unit intervals $AB_{k,t} = \{ab_{k,t-1}, \dots, ab_{k,t} - 1\}$; where $ab_{k,t}$ is the earliest interview date conducted by Afrobarometer round $t \in \{3, 4, 5, 6\}$ in cell k .³ Note that $ab_{k,2}$ is not present in the data, so that it is arbitrarily

3. When the interview date is missing – because the cell was not surveyed in that period – it is replaced by the country-specific average interview date, *i.e.* $|K_{p,t}|^{-1} \sum_{k \in K_{p,t}} ab_{k,t}$ where $K_{p,t}$ is the set of cells in country p surveyed during t .

set so as to impose a length of the first period approximately equal to the average of the others. Formally:

$$ab_{k,2} = ab_{k,3} - \frac{1}{3} \left(\frac{1}{K_6} \sum_{k \in K_6} ab_{k,6} - \frac{1}{K_3} \sum_{k \in K_3} ab_{k,3} \right)$$

where K_3 and K_6 are the sets of cells appearing in period 3 and 6, respectively.

The baseline measure of conflict is thus

$$C_{e,k,t} = \mathbb{1}\{\tilde{C}_{e,k,t} > 0\}$$

Where $\tilde{C}_{e,k,t}$ is the event-based definition of ethnic conflict. Formally:

$$\tilde{C}_{e,k,t} = \sum_{l \in k} \sum_{j \in AB_{k,t}} c_{e,l,j}$$

The third measure of conflict employed is the count of events with at least a conflict incident. Define $M_{k,t}$ as the set of unique months belonging to the interval $AB_{k,t}$, hence, the month-based measure of conflict is defined as:

$$\hat{C}_{e,k,t} = \sum_{m \in M_{k,t}} \mathbb{1}\left[\sum_{l \in k} \sum_{j \in m} c_{e,l,j} > 0\right]$$

Finally, the non-ethnic conflict variable employed at equation (2) is defined as:

$$Z_{-e,k,t} = \mathbb{1}\left[\sum_{l \in k} \sum_{j \in AB_{k,t}} c_{l,j} - \tilde{C}_{e,k,t} > 0\right]$$

Where $c_{l,j}$ is a conflict incident in location l at day j , which may or may not be related to a given ethnic group.

B Africa ethnolinguistic armed groups Dataset

This section introduces the *Africa ethnolinguistic armed groups* dataset, linking conflict actors to ethnolinguistic groups. This is the dataset used for the definition of ethnic conflict in the paper. Additional details on its construction and on the sources employed are available in the relevant codebook.

The dataset is a cross-section of armed actors active in sub-Saharan Africa, reporting information on their ethnolinguistic affiliation. In particular, a conflict actor is assigned to an ethnolinguistic group when the

former is organised along ethnic lines. Ethnic organisation primarily means that recruitment practices and actual composition of the armed actor overlap with ethnolinguistic cleavages. Occasionally, when the organisational structure was ambiguously defined, the assignment has also been made with reference to its conduct and targets. A note in the dataset explains the rationale behind each actor-ethnicity match. The list of armed actors comes from the Armed Conflict Location and Event Data (ACLED) Project. The cross-section contains all the non-state militias involved in conflicts taking place in various sub-Saharan countries over the first decade of the 21st century.

B.1 Dataset coverage

Spatially the dataset covers the eighteen countries present throughout Afrobarometer surveys rounds 3 to 6, as well as some neighbour areas. The countries included are: Benin, Botswana, Cape Verde, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Namibia, Nigeria, Senegal, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe. More precisely, countries are defined by the ensemble of grid cells intersecting the GADM polygon of the in-sample countries.^{4,5} Grid cells measure $0.5^\circ \times 0.5^\circ$, corresponding to roughly $55\text{km} \times 55\text{km}$ at the equator. Additionally, the dataset covers the conflict activity taking place in some of the cells bordering within-sample countries. In the specific, these are the cells falling within the 24-cell neighbourhood of an Afrobarometer respondent.⁶

The period considered depends on the interview dates of Afrobarometer surveys, with conflict activity analysed between Afrobarometer rounds “two” and six.⁷ Figure A1 gives a visual representation of the zones and

4. GADM is a portal providing administrative country divisions for the whole world. Access is free at the following link: <https://gadm.org/>

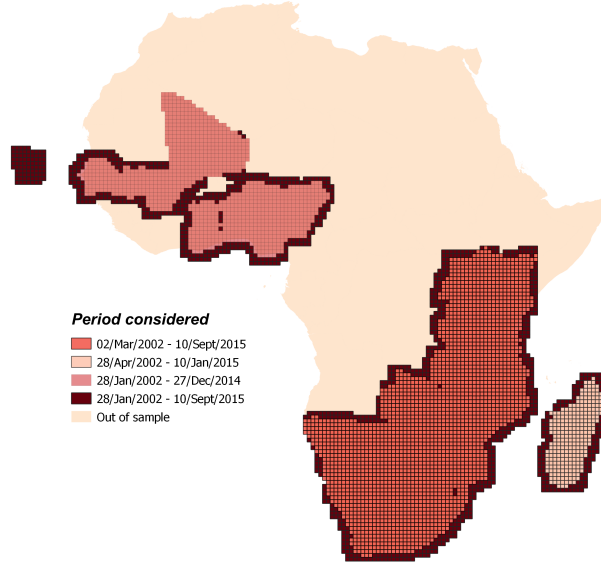
5. A cell belongs to a country when the two intersect. In case of cells spanning two or more within-sample countries, the cell was assigned to the country to which it belongs the most.

6. The X -cell neighbourhood of a given cell is the collection of X cells which surround it. Hence, the 24-cell neighbourhood of a cell c is the 5×5 square of cells centred around c and excluding it.

7. To be more precise, countries were divided into three macro-regions (*i.e.* West Africa, South-East Africa, and Madagascar), each having a different period in which conflict was considered. Formally, given a macro-region r , the upper bound of its period is given by $U_r = \max_{k \text{ in } r} ab_{k,6}$ where $ab_{k,6}$ is the earliest Afrobarometer 6 interview in cell k .

Similarly, the lower limit of the period is given by $L_r = \max_{k \text{ in } r} \tilde{a}b_{k,2}$ where $\tilde{a}b_{k,2}$ is the virtual

Figure A1: Dataset coverage



periods where ACLED conflict was taken into account.

Ethnolinguistic affiliation was possibly assigned to the following types of armed actors: rebel groups, political and identity militias. These are the actors identified by ACLED with interaction code 2, 3 and 4, respectively. Globally, they are involved in roughly 57% of the conflict incidents taking place across the 18 selected countries between May 2002 and November 2014. The dataset is muted, instead, on the possible ethnic identity of other ACLED actor categories. These are: state forces, rioters, protesters, civilians and other/external forces. The latter are an ephemeral actor category, appearing in very few conflict incidents related mostly to the French intervention in Mali (2012-2022) and various diplomatic events connected to international peacekeeping and monitoring operations. Rioters, protesters and civilians are, instead, involved in the vast majority of conflict incidents: jointly they appear in 69% of the conflict events considered. However, they are recorded by ACLED in general categories, such as “Protesters (Nigeria)” or “Civilians (Uganda)”.

interview date of Afrobarometer round 2. That is, inasmuch as $ab_{k,2}$ is not in the data, $\tilde{ab}_{k,2}$ is arbitrarily set so as to impose a length of the first period approximately equal to the average of the others, that is: $\tilde{ab}_{k,2} = ab_{k,3} - \frac{1}{3} \left(\frac{1}{K_6} \sum_{k \in K_6} ab_{k,6} - \frac{1}{K_3} \sum_{k \in K_3} ab_{k,3} \right)$, where K_3 and K_6 are the sets of cells appearing in period 3 and 6, respectively. The upper and lower interval limits of those cells not belonging to any within-sample country, are given by $\max_r U_r$ and $\min_r L_r$, respectively.

Clearly, this generic designation prevents the assignment of an ethnic identity. State forces are involved in as much as 34% of the conflict events. While being a significant actor in sub-Saharan Africa, the assignment of an ethnic affiliation is complicated by the multi-cultural nature of many African polities. Different state agencies have different ethnic allegiances, which moreover are prone to mutate over-time. Given these difficulties and the mole of research implied (there are 125 state actors active in the relevant sample), I preferred not to include them in the dataset, leaving this to future work.

B.2 Descriptive statistics

The dataset therefore has a good, if yet incomplete, coverage of sub-Saharan ethnic conflict throughout the 2000s and early 2010s. Table A2 provides summary statistics for ethnic violence, comparing it to all conflict activity (*All conflict*) and warfare involving any kind of non-state militias (*Non-ethnic & ethnic conflict*): between May 2002 and November 2014, across the 18 sub-Saharan countries sampled, the majority of violence did not follow ethnic lines. As shown in Figures A2a and A2b there is, however, significant heterogeneity across time and space. While ethnic conflict hovers around 250 events per quarter, other types of violence spiked after 2012 as a consequence of: political instability in Mali, the reignition of Boko Haram’s attacks in Nigeria, the intensification of al-Shabaab activity in Kenya, a small-lived resurgence of the Renamo rebel group in Mozambique, as well as a surge in riots and protests across both Ghana and South Africa.⁸ Spatially, instead, ethnic conflict tends to be concentrated in Nigeria, Uganda, and Kenya, with clusters also in Mali, Ghana, and Senegal. These countries experience also a lot of non-ethnic conflict activity, albeit other states, most notably South Africa and Zimbabwe, have equally massive presence of non-ethnic conflict.

Table A2: Conflict descriptive statistics (May 2002 - November 2014)

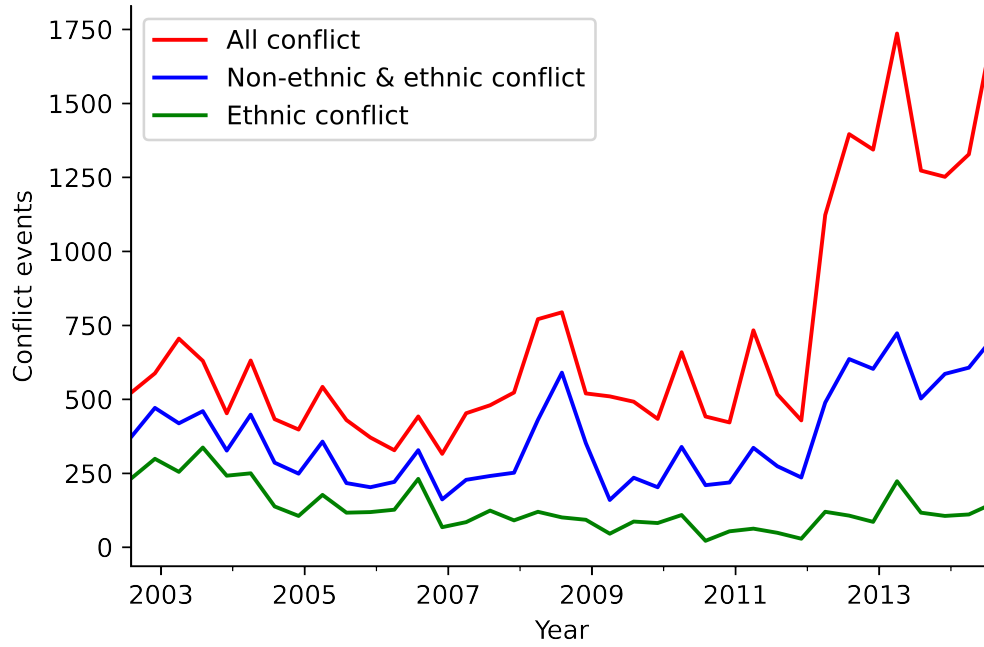
	All conflict	Non-ethnic & ethnic conflict	Ethnic conflict	Inter-ethnic conflict
<i>N</i> Events	27422	14093	4947	4771
<i>N</i> Cells	1401	1041	442	432
Fatalities	56832	49730	21405	20661

Notes: Data refer to the 18 sampled sub-Saharan countries.

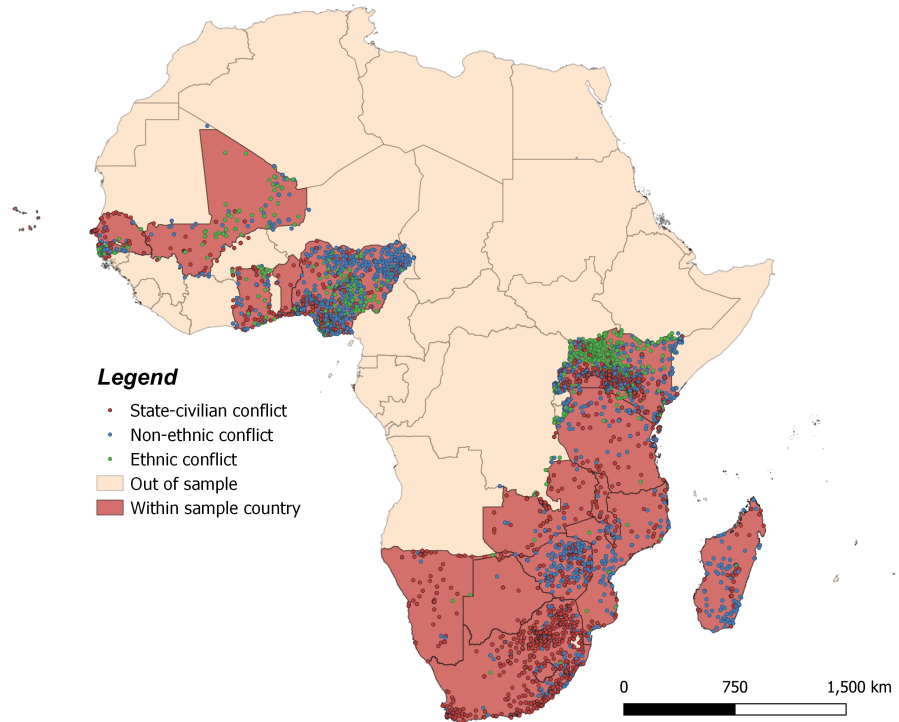
8. The UCDP website provides summary accounts of these conflicts.

Figure A2: Conflict trend and distribution (May 2002 - November 2014)

(a) Conflict events per quarter



(b) Conflict events across selected sub-Saharan countries



C Ethnocentrism

Ethnocentrism indicates the instinctive classification of the world into the superior ethnic 'us' and all the 'others'. As such, the concept refers to cultural mutable objects: peoples, nations, ethnicities.⁹ Yet, for how much artificial ethnicity is, ethnocentrism seems to be a fairly constant feature of human societies throughout the whole (pre-)history. It is particularly revealing to look at the names that many hunter-gatherer peoples used to refer to themselves. Consider the Eskimos of the Arctic zone, among which the Inuit are perhaps the most well-known ethnic group. All these societies called themselves by many names usually meaning the 'real people' (Gat 2006, pp. 50-51). Moving southwards in the New World, many American Indians also self-defined themselves as the sole legitimate humans. The native American tribes of Niimiipuu (Nez Perce), Lenape (Delaware), Lakota, Illiniwek referred invariably to themselves as 'men', 'people', or 'genuine Indians' (Barbero 2009). Similarly, Zhu/twasi, the ethnonym of a !Kung sub-group living across northern Botswana and Namibia, is a self-appellation translating as 'genuine people' (Lee 1979, p. 38). Even more telling is the case of the Yanomamo horticulturalists, still inhabiting the fluvial zone between Brazil and Venezuela. "Yanomamo in fact means 'humanity', or at least the most important segment of humanity. All other peoples are known by the term *naba*, a concept that implies an invidious distinction between 'true' man and 'sub-human' man."¹⁰ These examples are symptomatic of a broader pattern that extends to many other tribal societies, including Aboriginal Australian groups (Hamilton 2006, footnote 8).

If foraging peoples tended to deny the very humanity of other ethnic groups by considering themselves as the only 'legitimate people', the picture did not change much with the rise of states. As it is well known, ancient Greeks pejoratively defined all the rest of the world as barbarians. The very term barbarian comes from the 'bar-bar' sound whereby Greeks mimicked what they considered grotesque and incomprehensible non-Greek languages. In the classical world, the term will be later enlarged to indicate all those people living outside of the Roman Empire. But this was not a Mediterranean peculiarity. From Chinese to Persians and Indians, much of the pre-modern states have applied the derogative barbarian category to the outer world (Scott 2009 & 2017).

9. For some historical examples on the malleability of ethnic identities see: Barbero (2009), Pohl (2015), Tilly (1994), and Welsh (1996).

10. Napoleon Chagnon (1968), *Yanomamo social organization and warfare*, Ann Arbor (MI): University Microfilms, as quoted in Gat (2006, p. 51)

As the brief discussion above has shown, the ethnic-based ‘us-them’ demarcation of the world is deeply in-grained in human psyche. The overwhelming force of ethnocentrism, as well as its omnipresence in human societies, are difficult to explain with exclusive reference to social factors. Socio-biology is, instead, on a much firmer foot to explain the ubiquitousness of this phenomenon. Various evolutionary theories have been advanced to explain humans’ peculiarly strong in-group bias.¹¹ The most widely accepted one understands ethnocentrism as residual, almost a manipulation, of natural selection on pro-kinship behaviour.¹²

From an evolutionary perspective, purely altruistic behaviour – diminishing one’s own biological fitness at no personal benefit – can naturally emerge only towards relatives. The cold mechanics of natural selection plainly imply that a given behaviour can be naturally selected if the genes mandating for it are passed on to the next generation. Genetic material is transmitted both through offspring and kin mates. Relatives share some of their genes among themselves, in diminishing proportion as the kin distance increases.¹³ Humans, as much as other animals,¹⁴ have thus developed a natural inclination to help their relatives, with greater sacrifices easily exerted for closer kin.

11. See Dennen (2000) and Wrangham (2019, ch. 7) for a review of some of these theories. The most popular among economists is the so-called parochial altruism theory (Bernhard et al. 2006; Bowles 2008, 2009, 2012; Choi and Bowles 2007), whereby in-group altruism emerged because it gave a competitive edge in conflict among hunter-gatherer bands: more cohesive groups could subjugate the others, hence the natural selection of cooperative traits. For how appealing and elegant the argument is, it rests on the controversial notion of group-selection (Pinker 2012); does not fit well with the raid-and-ambush nature of prehistoric warfare (Gat 2006); can not be extended to other animal species characterised by intensive group-conflict (Wrangham 2019, ch. 7); implies a close relationship between in-group favouritism and out-group aggressiveness, which is not however always observed (Brewer 2007; Yamagishi and Mifune 2016). More robust are the arguments connecting human altruism to a millennial process of self-domestication (Wrangham 2019). These theses are, however, not antithetical to the one here illustrated; they are, indeed, probably complementary inasmuch as they deal with generic human cooperation, rather than limited to co-ethnics.

12. The following discussion relies heavily on the treatment in Gat (2006), particularly at chapters 3 and 7.

13. For example, siblings share, on average, 50% of their genes, while cousins just 12.5%.

14. The case of bees is emblematic. They live in colonies whose vast majority of members are siblings or even clones. It is not surprise then to observe incredibly altruistic behaviours: individual bees are ready to sacrifice their lives for the survival of the whole colony, which is a huge repository of the genes present in a single bee (Gat 2006, p. 44).

As a species, *Homo Sapiens* developed almost uniquely during the pre-agricultural era, when humans roamed the world as small bands of hunter-gatherers. For most of *Sapiens* existence, people lived in extended family clans numbering between 20 and 70 members (Kelly 2007). Developments in linguistic skills around perhaps 70.000 years ago, allowed the formation of tribes, regional groups reuniting several familiar clans (Gat 2006, Harari 2014). The tribe, numbering in the hundreds, shared a common language and was characterised by extensive marriage exchanges among its constituent clans. Therefore, in these prehistoric societies kinship and culture overlapped to a great extent. Beyond phenotypic traits, common language became a primary cue for kinship relations. While not all members of these regional groups were close relatives, they had far greater genetic proximity among themselves than with respect to foreigners. Language was effectively a good indicator of kinship.

Moreover, as these regional groups developed particular cultural attributes, kin-mandated cooperation was further strengthened in virtue of common cultural practices, most notably a shared language. In short, regional hunter-gatherer groups were “bound together by mutually reinforcing and overlapping ties of kinship, social co-operation, and cultural distinctiveness” (Gat 2006, p. 50).

This is the socio-biological essence of ethnocentrism. In the pre-agricultural world, where ethnicity predicted kinship, it is easy to see the evolutionary rationale of the us-them discrimination. Yet, as human societies grew larger and more complex, the application of kin-solidarity to culturally related people (*i.e.* co-ethnics) definitely lost any evolutionary aim. When during the Second World War Japanese kamikazes committed suicide against US ships, they were neither increasing their biological fitness, nor that of any distant relative. Nonetheless, some of them were volunteers, enthusiastically going to risk their lives for the *motherland*. This seemingly irrational behaviour shall be read as a manipulation, an extension beyond its original objective, of kin-based solidarity within small groups.¹⁵ It is this the atavistic component of contemporary xenophobia, the same often perplexing observers of ethnic cleansing and genocides.

15. The persistence of naturally selected behaviours even if devoid of their evolutionary *raison d'être*, albeit surprising, is typical of humankind. For example, in the same logic, as humans we have a natural love for sweet food, even though sweetness does not serve any more its original evolutionary aim of signalling fruit maturity.

D Baseline - Robustness tests

This section reports additional robustness tests of the baseline models estimated in Section 4 of the paper. The estimating equation of interest is:

$$y_{i,e,k,t} = \mu_{k,t} + \lambda_{e,t} + \beta C_{e,k,t} + \delta' x_{i,e,k,t} + u_{i,e,k,t} \quad (1)$$

Where all the variables have the same meaning as in the paper. The coefficient of interest is the marginal impact of ethnic conflict (β), which is expected to be positive.

D.1 Outlier observations

As clear from Figure 2 in the paper, ethnic conflict is spread very unevenly across sub-Saharan Africa. For each country, Figure A3 reports the cell-level average of conflict events.¹⁶ Kenya is the country with the highest density of ethnic conflict events, primarily reflecting the violent nature of its ethnic-based political system, as well as conflict spill-overs stemming from its troubled Somali and Ugandan neighbours. On the other end of the spectrum are Cape Vert, Lesotho, and Madagascar, which did not experience any conflict event during the whole period. As shown in Figure A4, baseline results are overall robust to the exclusion of each country from the sample.

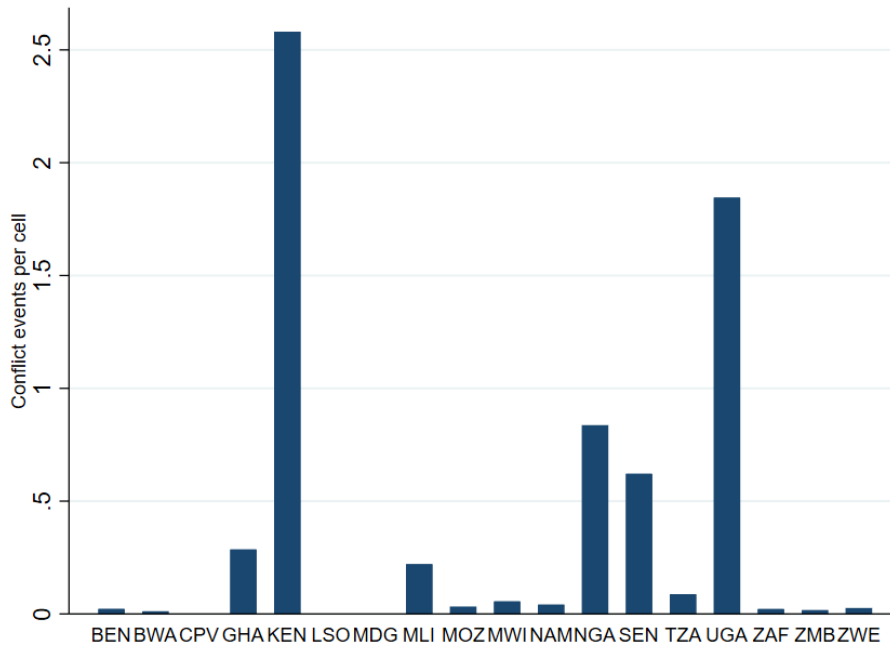
A related concern is that the impact of conflict on social cohesion might be driven by few influential ethnic groups. Table A3 reports the total number of conflict events for the 64 ethnolinguistic groups that have experienced at least one conflict incident. There is, indeed, some substantial heterogeneity in warfare exposure. Kikuyu, Ijaw, Diola and Luo are the four most war-prone groups, the only experiencing at least 40 conflict incidents. The Kikuyu and Luo of Kenya have been exposed to different types of conflict, ranging from tribal clashes over cattle and land, to urban disorders in the slums of Nairobi at the hands of the Kikuyu-led 'Mungiki Militia' and the Luo-controlled 'Taliban Militia'. The Ijaw of Nigeria have taken part in conflicts interesting the oil-rich Delta

16. That is, the event-based definition of conflict is first aggregated at the country level and then divided by the number of cells in each country. Formally:

$$\frac{1}{|\mathcal{K}_p|} \sum_{k \in \mathcal{K}_p} \tilde{C}_{e,k,t}$$

where \mathcal{K}_p is the set of cells in country p and $\tilde{C}_{e,k,t}$ is the event-based definition of conflict defined in the data section of the paper.

Figure A3: Ethnic conflict in selected sub-Saharan countries



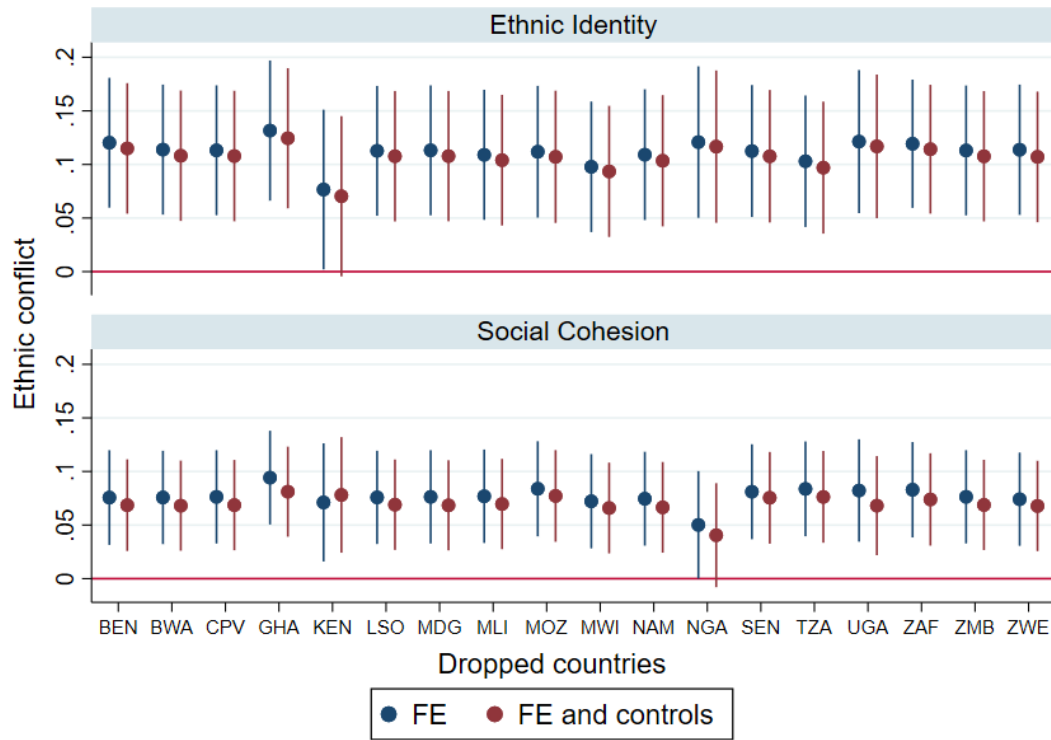
Notes: Conflict events are aggregated at the country-level throughout the whole sample period.

region of the Niger river. The area has known “decades of environmental pollution, economic underdevelopment and political marginalization”(Asuni 2009, p. 3), which have ultimately resulted in fertile ground for inter-ethnic tensions. Hence, the involvement of the Ijaw in various ethnic militias and loose armed groups, such as the ‘MEND: Movement for the Emancipation of the Niger Delta’ (Hazen and Horner 2007). The Diola of Senegal have been at the forefront of the independence struggle for the Casamance region, taking the lead in rebel groups such as the ‘MFDC: Movement of Democratic Forces of Casamance’.¹⁷

The baseline model is thus re-estimated excluding these outlier groups. Results are reported in Table A4. The first two columns substantiate the robustness of results to the exclusion of Kikuyu respondents from the sample. Columns three and four run a more demanding specification whereby also Ijaw, Diola and Luo are excluded, limiting the sample to ethnolinguistic groups reporting less than 40 conflict incidents. Marginal impacts remain positive and significant. Finally, in the last two columns, I drop all the ethnolinguistic groups not experiencing any

17. For an overview of the conflict and the role of MFDC see: <https://ucdp.uu.se/conflict/375> and <https://ucdp.uu.se/actor/529>.

Figure A4: OLS regressions of Ethnic Identity & Social Cohesion on Ethnic conflict – Exclude countries



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic conflict from linear regressions sequentially excluding one country at time. Fixed effects correspond to cell×period and ethnic×period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. Standard errors are clustered at the cell-ethnic-period level.

conflict activity. Even though the sample size is roughly halved, the estimated coefficients are positive and always statistically different than zero.

Table A3: Conflict incidence by ethnolinguistic group

ETHNOLINGUISTIC GROUP	CONFLICT INCIDENCE	ETHNOLINGUISTIC GROUP	CONFLICT INCIDENCE
Acholi	26	Lozi	1
Afrikaans	1	Lugwere	1
Akan	4	Luhya	3
Alur	3	Lukhonjo	5
Ateso	13	Lumasaaba	1
Chewa	2	Luo	40
Dagbani	3	Lusoga	1
Diola	40	Mampruli	4
Edo	1	Ndebele	1
Eggon	3	Ngakarimajong	11
Ewe	6	Ogoni	3
Fulani	16	Oroma	32
Hausa	3	Oshiwambo	2
Herero	2	Pokot	13
Idoma	1	Pulaar/Toucouleur	1
Igbo	25	Rufumbira	36
Ijaw	81	Rukiga	7
Isoko	1	Rukwangali	1
Japadhola	2	Runyankole	1
Kalabari	1	Runyoro	27
Kalenjin	32	Shangaan/Tsonga	7
Kanuri	1	Shona	1
Karanga	1	Somali	9
Kikuyu	88	Sotho	1
Kimasai	8	Tamasheq	30
Kimeru	5	Tarok	6
Kisii	3	Tiv	3
Konkomba	2	Tonga	1
Kupsabinyi	1	Turkana	5
Kuria	12	Urhobo	4
Kusal	5	Xhosa	2
Langi	1	Yoruba	36

Notes: Conflict incidence refers to number of conflict events experienced by an ethnolinguistic group throughout the whole period of analysis. Only ethnolinguistic groups experiencing at least a conflict event are included in the table.

Table A4: OLS regressions of Ethnic Identity & Social Cohesion on Ethnic conflict – Outlier ethnic groups

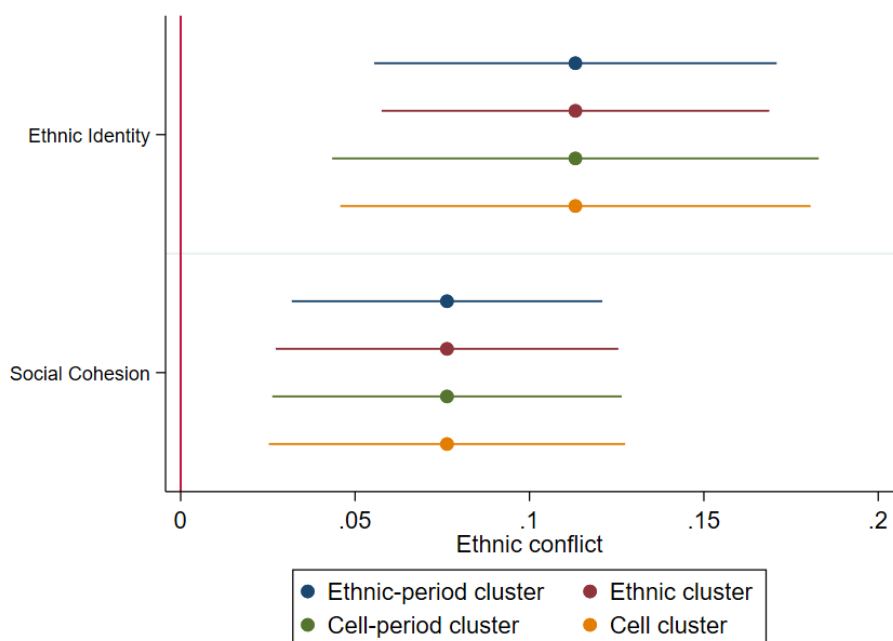
	(1)	(2)	(3)	(4)	(5)	(6)
ETHNIC IDENTITY						
Ethnic conflict	0.094 (0.041)**	0.088 (0.041)**	0.121 (0.043)***	0.118 (0.044)***	0.096 (0.042)**	0.094 (0.042)**
R^2	0.22	0.22	0.22	0.22	0.21	0.22
N	102,191	99,990	100,251	98,059	49,955	49,429
SOCIAL COHESION						
Ethnic conflict	0.071 (0.028)**	0.072 (0.028)**	0.070 (0.032)**	0.072 (0.031)**	0.087 (0.031)***	0.079 (0.031)**
R^2	0.23	0.27	0.23	0.27	0.23	0.29
N	107,979	105,711	106,058	103,798	51,498	50,960
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual cov.	No	Yes	No	Yes	No	Yes
Drop Kikuyu	Yes	Yes	Yes	Yes	No	No
Drop Ijaw, Diola & Luo	No	No	Yes	Yes	No	No
Drop peaceful groups	No	No	No	No	Yes	Yes

Notes: The sample includes Afrobarometer rounds 3 to 6. Fixed effects correspond to cell \times period and ethnic \times period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. The *Ethnic conflict* dummy signals the presence of conflict events involving the ethnolinguistic group of an observation and taking place in her cell of residence during the relevant period. *Ethnic Identity* captures the saliency of ethnic identification, while *Social Cohesion* is an index combining three dimensions of pro-sociality: attendance at community meetings, participation in instances of collective action, and membership in community associations. Standard errors are clustered at the cell-ethnic-period level and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels

D.2 Standard error clustering

In the baseline specification standard errors are allowed to be correlated at the cell-ethnolinguistic-period level. The level of clustering was chosen so as to match the level of assignment of the conflict indicator. I here relax this assumption and allow standard errors to be correlated within each: ethnolinguistic-period pair, ethnolinguistic group, cell-period pair, and within each $0.5^\circ \times 0.5^\circ$ cell. As shown in Figure A5, sign and statistical significance of the baseline estimates are always reproduced: the validity of the results does not hinge upon the level of clustering.

Figure A5: OLS regressions of Social identity and cohesion on Ethnic conflict – Alternative standard error clustering



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic conflict under different levels of standard error clustering. Fixed effects correspond to cell×period and ethnic×period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator.

D.3 Alternative treatment and outcome specifications

This section investigates robustness to alternative specifications of ethnic conflict (Table A5) and social cohesion (Table A6).

The first two columns of Table A5 check whether results are sensitive to possible measurement errors in the recording of conflict. Time-exact conflict thus considers only events whose date is identified with certainty by ACLED: baseline results remain unchanged.¹⁸ The successive pair of columns addresses concerns related to intra-ethnic conflict. When violence cuts through an ethnolinguistic group, the issue of contention is likely unrelated to ethnicity, making harder for ethnocentric feelings to materialise. Indeed, as shown in columns 3-4, when excluding these types of violent events, point estimates are greater than in the baseline models. The last two columns of Table A5, instead, check robustness to the inclusion of conflict events involving state forces. While this might make omitted variable issues more compelling – if anything because

18. Formally, time-exact conflict consider only events classified by ACLED with a time precision equal to 1.

state involvement in conflict is more directly related to changes in overall state capacity and legitimacy –, results are robust to this inclusion.

Table A6 reports estimates of models employing the components of the social cohesion index as dependent variables. The impact of ethnic conflict on the three measures of social cohesion mimics that on the aggregated index: marginal effects are positive and become significant as cell-period fixed effects are included (columns 2-4).

Table A5: OLS regressions of Social identity and cohesion on Ethnic conflict – Time-precise, external and state conflict

	(1)	(2)	(3)	(4)	(5)	(6)
ETHNIC IDENTITY						
Time-exact conflict	0.094 (0.039)**	0.088 (0.039)**				
External conflict			0.125 (0.038)***	0.121 (0.038)***		
State conflict					0.086 (0.036)**	0.083 (0.036)**
R^2	0.22	0.22	0.22	0.22	0.22	0.22
N	103,588	101,378	103,588	101,378	103,588	101,378
SOCIAL COHESION						
Time-exact conflict	0.091 (0.027)***	0.080 (0.026)***				
External conflict			0.087 (0.028)***	0.077 (0.027)***		
State conflict					0.069 (0.025)***	0.061 (0.024)**
R^2	0.22	0.27	0.22	0.27	0.22	0.27
N	109,371	107,094	109,371	107,094	109,371	107,094
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual cov.	No	Yes	No	Yes	No	Yes

Notes: The sample includes Afrobarometer rounds 3 to 6. Fixed effects correspond to cell×period and ethnic×period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. *Time-exact conflict* includes only incidents of ethnic violence whose date is exactly identified by ACLED; *External conflict* is defined as conflict not involving two armed groups belonging to the same ethnolinguistic cluster; and *State conflict* includes also clashes between state forces and ethnic groups. *Ethnic Identity* captures the saliency of ethnic identification, while *Social Cohesion* is an index combining three dimensions of pro-sociality: attendance at community meetings, participation in instances of collective action, and membership in community associations. Standard errors are clustered at the cell-ethnic-period level and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table A6: OLS regressions of Social cohesion components on Ethnic conflict

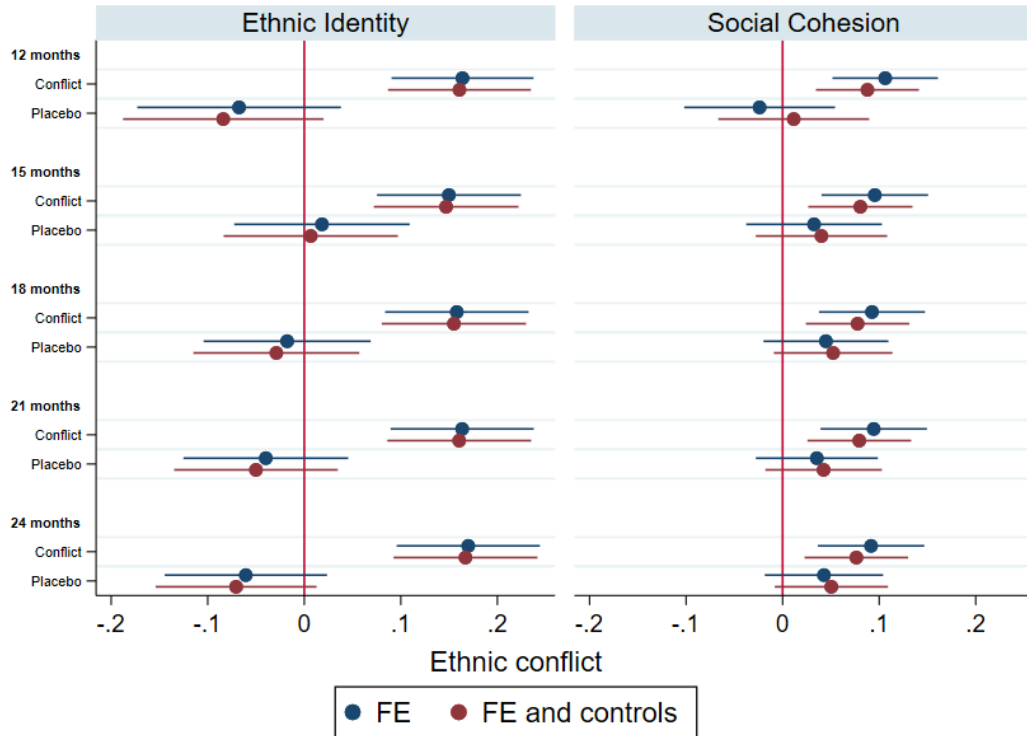
	(1)	(2)	(3)	(4)
ATTENDANCE COMMUNITY MEETINGS				
Ethnic conflict	0.042 (0.031)	0.099 (0.040)**	0.118 (0.040)***	0.099 (0.039)**
R^2	0.16	0.22	0.22	0.27
N	110,763	110,735	110,718	108,389
COLLECTIVE ACTION				
Ethnic conflict	0.020 (0.053)	0.097 (0.039)**	0.098 (0.040)**	0.089 (0.039)**
R^2	0.15	0.23	0.23	0.26
N	110,511	110,482	110,464	108,151
MEMBERSHIP COMMUNITY ASSOCIATIONS				
Ethnic conflict	0.045 (0.030)	0.067 (0.032)**	0.063 (0.033)*	0.062 (0.033)*
R^2	0.11	0.16	0.17	0.18
N	110,378	110,352	110,335	108,029
Period FE	Yes	No	No	No
Cell FE	Yes	No	No	No
Ethnic FE	Yes	Yes	No	No
Cell-period FE.	No	Yes	Yes	Yes
Ethnic-period FE.	No	No	Yes	Yes
Individual cov.	No	No	No	Yes

Notes: The sample includes Afrobarometer rounds 3 to 6. Individual controls include: age and its square, gender, and an urban-rural indicator. The *Ethnic conflict* dummy signals the presence of conflict events involving the ethnolinguistic group of an observation and taking place in her cell of residence during the relevant period. The dependent variables are the components of the Social Cohesion index used in the baseline analysis. *Attendance Community Meetings* measures participation in community associations, *Collective Action* captures involvement in instances of collective action, *Membership Community Associations* measures the degree of engagement in associations with a communitarian scope. Standard errors are clustered at the cell-ethnic-period level and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

D.4 Placebo

To strengthen a causal interpretation of the results, Figure A6 reports estimates of the baseline model as augmented with an indicator function for conflict taking place $X \in \{12, 15, 18, 24\}$ months after the relevant

Figure A6: OLS regressions of Social identity and cohesion on Ethnic conflict – Future conflict activity (Placebo)



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic conflict before (*Conflict*) and after (*Placebo*) the relevant Afrobarometer interview. Fixed effects correspond to cell×period and ethnic×period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. Standard errors are clustered at the cell-ethnic-period level.

Afrobarometer interview rounds 3 to 5.¹⁹ Placebo coefficients are always statistically indistinguishable from zero and smaller than the actual conflict dummy. Point estimates on the baseline conflict measure remain instead positive and significant.

D.5 Migration

This last subsection addresses concerns related to the possible confounding effect of migration. On the one hand, population pressures

19. Formally, while the baseline conflict dummy consider events prior to $ab_{k,t}$, the earliest interview date of Afrobarometer round $t \in \{3, 4, 5, 6\}$ in cell k ; placebo conflict is aggregated in the X months successive $ab_{k,\tau}^{late}$, the latest interview date of Afrobarometer round $\tau \in \{3, 4, 5\}$ in cell k .

can affect the likelihood of conflict.²⁰ On the other hand, warfare often results into massive displacement of people.²¹ If these migratory patterns correlate with social cohesion, there are sample selection issues that can invalidate inference. As it has also been shown empirically in the case of the Nepalese “People’s war” (1996-2006), less pro-social people might be more likely to flee away from their communities at times of conflict (Gilligan et al. 2014). If this is the case, migration, not ethnocentrism, might actually explain the positive impact of warfare on social cohesion here found. Yet, in the present context this seems unlikely. First, cell-time fixed effects should account for migratory patterns that are not specific to some given ethnic groups. Second, most of the ethnic violence observed in the sample concerns low-intensity conflicts, which are unlikely to trigger mass-scale migration.

In order to give empirical support to these intuitions, I analyse the effect of ethnic conflict over changes in ethnic population shares. I thus collapse the sample at the ethnolinguistic-cell-period level and drop all the ethnic-cell pairs not appearing throughout the four periods. Hence, I define ethnic share change as $\Delta Etn_share_{e,k,t} = \pi_{e,k,t} - \pi_{e,k,t-1}$, where $\pi_{e,k,t}$ is the share of respondents belonging to ethnolinguistic group e in cell-period $k-t$. If ethnic conflict does not result into migration, we would expect the change in these shares to be unaffected by warfare exposure. Table A7 performs a t-test for difference in means across the two sub-samples determined by the ethnic conflict dummy. The alternative hypothesis that the mean change is different across the two groups is rejected.

If conflict does not affect the ethnic composition of the various cell-period units, it might still be the case that it changes the internal composition of ethnic groups. In particular, conflict might lead to gender-specific migration, with women disproportionately leaving war-torn areas. The different gender balance might then influence the social cohesion dynamics at the local level. I thus analyse changes in male ratios, which are defined as $\Delta Male_ratio_{e,k,t} = Male_ratio_{e,k,t} - Male_ratio_{e,k,t-1}$, where $Male_ratio_{e,k,t} = \#male_respondents_{e,k,t} / \#respondents_{e,k,t}$. The second row of Table A7 substantiates that gender composition is relatively unaffected by conflict: while male ratios are slightly higher among ethnic groups

20. For example, Hendrix and Salehyan (2012) and Seter (2016) discuss how weather calamities can result into migration towards areas spared by the climatic disaster. The resulting population pressures can then put a strain on local resources, ultimately leading to open conflict.

21. According to the World Bank, in 2016 across sub-Saharan Africa, there were more than 12 million internally displaced people for reasons related to conflict and violence.

Table A7: T-test for differences in population share and male ratio changes across peaceful and violent ethnolinguistic groups

	Mean Peace	Mean War	p-value	<i>N</i> peace	<i>N</i> war
Δ Ethnic share	0.00	-0.01	0.47	2265	81
Δ Male ratio	0.00	0.03	0.20	2265	81
Ethnolinguistic groups	130	33	.	3019	109
Cells	467	61	.	3019	109
Periods	4	4	.	3019	109

Notes: The sample is a balanced panel with each ethnolinguistic-cell pair observed in all Afrobarometer rounds 3 to 6.

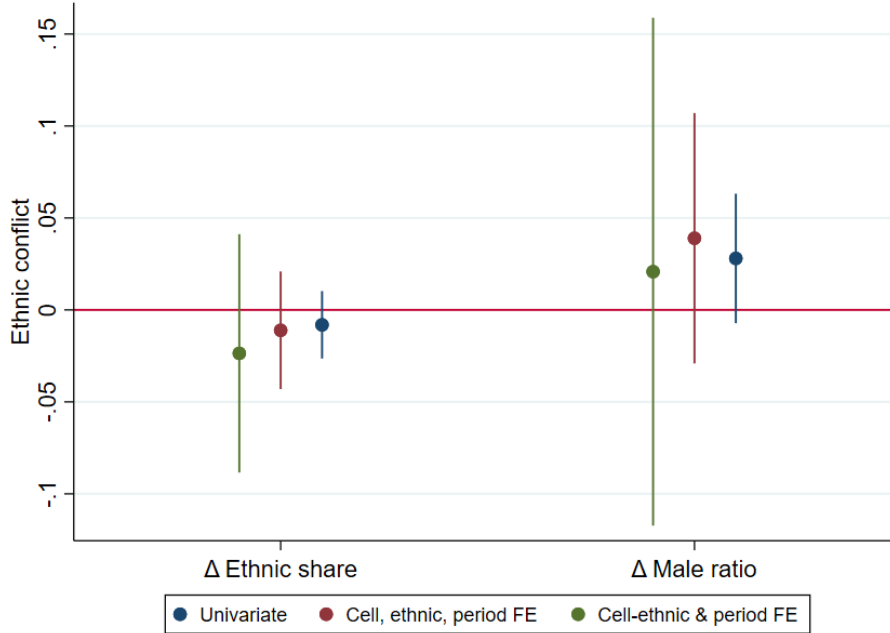
exposed to conflict, this difference is statistically insignificant.

Further empirical support to the average differences shown at Table A7, is given in Figure A7, reporting point estimates and 90% confidence intervals from the regressions:

$$\Delta Mig_{e,k,t} = \alpha + \theta_{e,k} + \Phi_t + \gamma C_{e,k,t} + \varepsilon_{e,k,t}$$

Where: $\Delta Mig_{e,k,t} \in \{\Delta Etn_share_{e,k,t}, \Delta Male_ratio_{e,k,t}\}$, $\theta_{e,k}$ are cell-ethnic fixed effects, Φ_t are period dummies, and $\varepsilon_{e,k,t}$ are robust standard errors. All the estimated coefficients are small and statistically indistinguishable from zero.

Figure A7: OLS regressions of Ethnic share and Male ratio changes on Ethnic conflict



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic conflict. Fixed effects correspond to cell×period and ethnic×period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. Standard errors are allowed to be heteroskedastic. The sample is a balanced panel of cell-ethnic units.

E Mechanisms - Robustness tests

E.1 Distant conflict

The model investigating the impact of remote conflict activity on social identity and cohesion is:

$$y_{i,e,k,t} = \mu_{k,t} + \lambda_{e,t} + \phi \mathcal{C}_{e,n(k),t} + \delta' x_{i,e,k,t} + u_{i,e,k,t}$$

Where $\mathcal{C}_{e,n(k),t}$ is the ethnic conflict indicator computed over the eight-cell neighbourhood of cell k . The meaning of the other variables is the same as in the paper. The coefficient of interest is ϕ , which is expected to be positive.

Figure A8 checks robustness to: (i) alternative conflict measures, namely, the count of conflict events and of months with at least one event; (ii) alternative actor-ethnicity match procedures, namely, that excluding imprecise matches and that dropping matches involving subgroups of Afrobarometer languages; (iii) alternative samples, namely,

that dropping ethnolinguistic groups not present in all Afrobarometer surveys and that using the Afrobarometer question on ethnicity as marker defining ethnolinguistic groups; (iv) a binary recoding of the outcome variables; (v) the inclusion of further individual-level covariates, namely, two sets of socio-economic and political controls. Results are overall robust, albeit the impact of distant conflict on social cohesion becomes at time marginally insignificant.

The last panel (*Remote conflict*) of Figure A8 explores alternative distances over which external conflict is aggregated. In particular, I consider ethnic conflict aggregated over the 24-cell neighbourhood of each observation. This measure includes conflict activity taking place very far away from Afrobarometer respondents (up to 233km). Unsurprisingly coefficients are smaller and statistical significance is not always achieved. Overall, however, the consistency of the sign is indicative that identity dynamics are at play: it is hard to see how conflict can directly affect security and economic incentives when taking place over 200km away.

E.2 Ethnic vs non-ethnic conflict

The model comparing ethnic and non-ethnic conflict is the following:

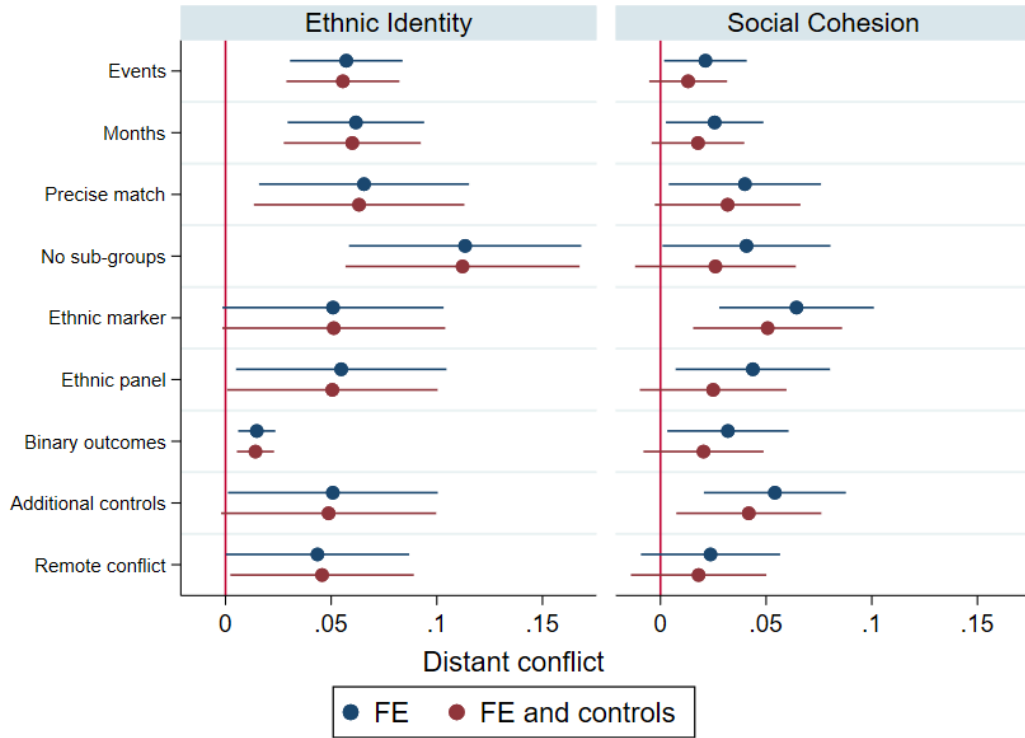
$$y_{i,e,k,t} = \mu_{k,t} + \lambda_{e,t} + \beta C_{e,k,t} + \psi Z_{-e,k,t} + \delta' x_{i,e,k,t} + u_{i,e,k,t} \quad (2)$$

Where all the variables have the same meaning as in the paper. The coefficients of interest are β and ψ , whereby we expect the former to be greater than the latter.

Figure A9 checks robustness to: alternative actor-ethnicity match procedures, sample-wide restrictions, binary recording of outcome variables, and the inclusion of additional controls.²² All the regressions include the usual set of fixed effects and individual-level covariates. The point estimates for ethnic conflict (β) are always positive and, with few exceptions, statistically significant. The coefficient on non-ethnic conflict is instead mostly negative and statistically indistinguishable from zero.

22. Note that we can not check robustness to the event-based definition of conflict inasmuch as ethnic conflict events and non-ethnic conflict events are collinear to the cell-period fixed effects.

Figure A8: OLS regressions of Social identity and cohesion on Distant ethnic conflict – Robustness tests



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of distant ethnic conflict. Fixed effects correspond to cell \times period and ethnic \times period dummies. Individual controls include: age and its square, gender, and an urban-rural indicator. *Distant conflict* is spatially aggregated on the 8-cell neighbourhood of each observation. *Remote conflict* is instead spatially aggregated on the 24-cell neighbourhood of each observation. In the eight panel (*Additional controls*), regressions labelled *FE* comprehend fixed effects, individual and socio-economic controls; regressions labelled *FE and controls* further add political covariates. Standard errors are clustered at the cell-ethnic-period level.

E.3 Heterogeneous effects

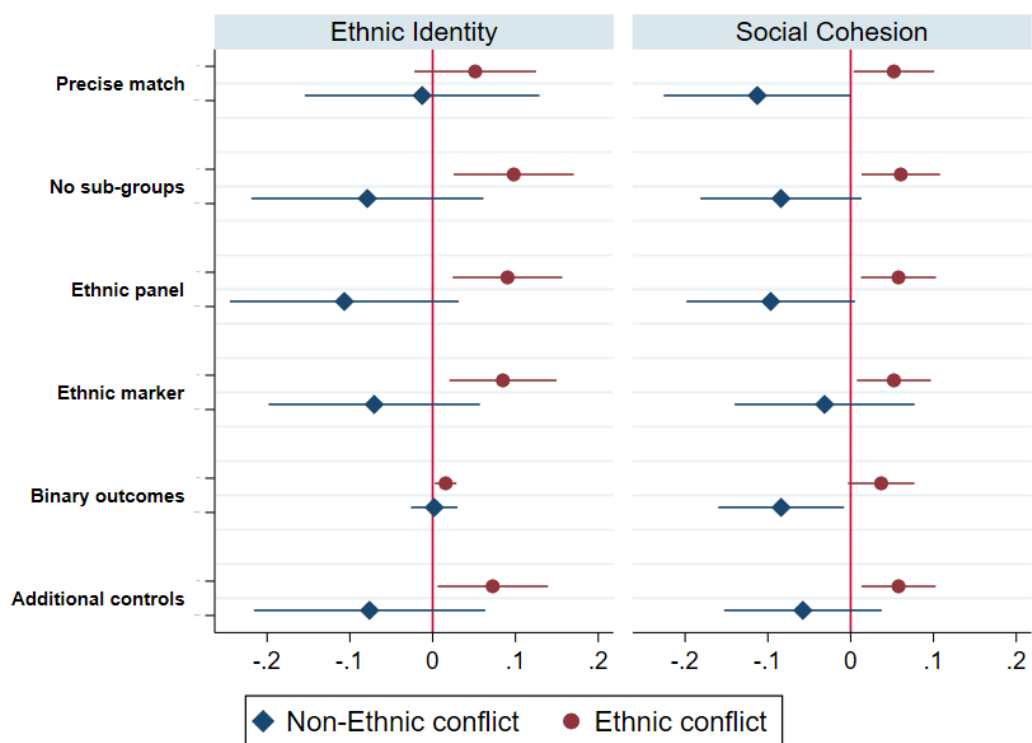
The model investigating heterogeneous effects with respect to cell-period ethnic polarisation is:

$$y_{i,e,k,t} = \mu_{k,t} + \lambda_{e,t} + \beta C_{e,k,t} + \alpha(C_{e,k,t} \times ELP_{k,t}) + \delta' x_{i,e,k,t} + u_{i,e,k,t} \quad (3)$$

Where all the variables have the same meaning as in the paper. Interest lies in the joint effect of conflict and ethnolinguistic polarisation (ELP): the parameter α , which we expect to be negative.

Figure A10 checks robustness to: alternative conflict measures, alternative actor-ethnicity match procedures, sample-wide restrictions, binary recording of outcome variables, and the inclusion of additional

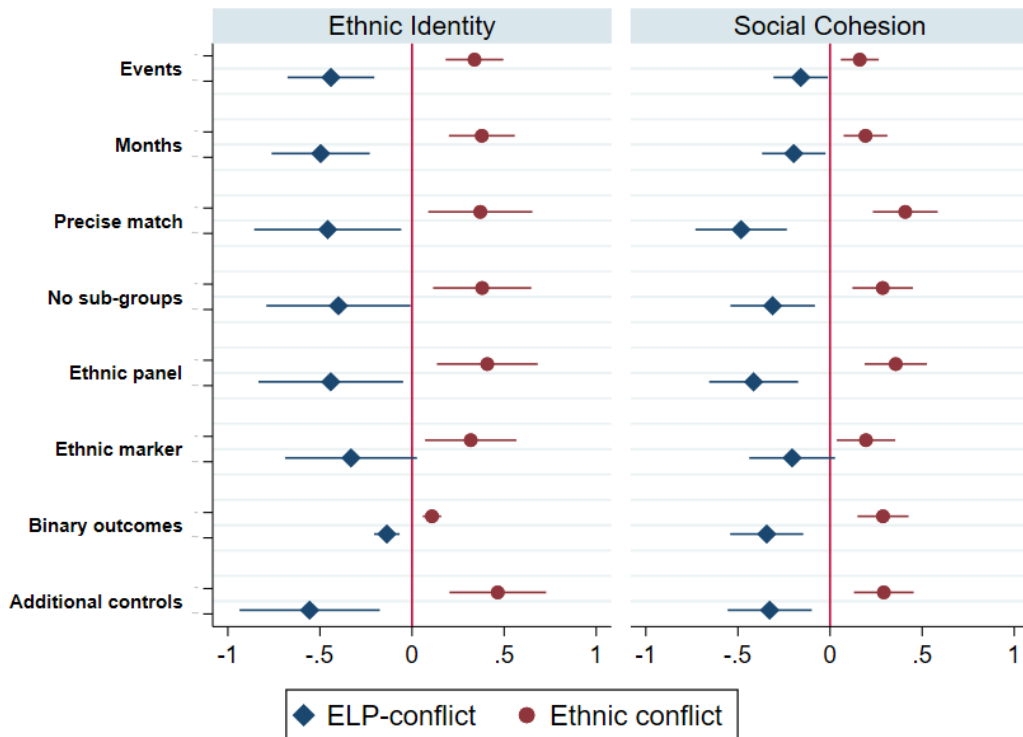
Figure A9: OLS regressions of Social identity and cohesion on Ethnic and non-ethnic conflict events – Robustness tests



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic and non-ethnic conflict. All regressions include: cell \times period and ethnic \times period fixed effects, as well as individual controls for age and its square, gender, and urban residence. In the last panel (*Additional controls*), socio-economic and political covariates are added. Standard errors are clustered at the cell-ethnic-period level.

controls. All the regressions include the usual set of fixed effects and individual-level covariates. Apart for the regressions employing the alternative marker for ethnolinguistic groups, the ELP-conflict interaction is always significantly negative.

Figure A10: Heterogeneous effects by ELP – Robustness tests



Notes: The figure reports point estimates, alongside their 90% confidence intervals, of ethnic conflict and its interaction with the index of ethnolinguistic polarisation. All regressions include: cell×period and ethnic×period fixed effects, as well as individual controls for age and its square, gender, and urban residence. In the last panel (*Additional controls*), socio-economic and political covariates are added. Standard errors are clustered at the cell-ethnic-period level.