Mineral resources and the salience of ethnic identities*

Nicolas Berman† Mathieu Couttenier‡ Victoire Girard§

This version: July 14, 2022

Abstract. This paper shows how ethnic identities may become more salient due to natural resources extraction. We combine individual data on the strength of ethnic – relative to national – identities with geo-localized information on the contours of ethnic homelands and on the timing and location of mineral resources exploitation in 25 African countries, from 2005 to 2015. Our strategy takes advantage of several dimensions of exposure to resources exploitation: time, spatial proximity, and ethnic proximity. We find that the strength of an ethnic group identity increases when mineral resource exploitation in that group's historical homeland intensifies. We argue that this result is at least partly rooted in feelings of relative deprivation associated with the exploitation of the resources. We show that such exploitation has limited positive economic spillovers, especially for members of the indigenous ethnic group; and that the link between mineral resources and the salience of ethnic identities is reinforced among members of powerless ethnic groups, and groups with strong baseline identity feelings or living in poorer areas, or areas with a history of conflict. Put together, these finding suggest a new dimension of the natural resource curse: the fragmentation of identities, between ethnic groups and nations.

JEL Codes: J15, N57, 012, O55, Q32.
Keywords: identity, ethnicity, natural resources.

*We thank the editor, Ekaterina Zhuravskaya and three anonymous referees for their thoughtful suggestions. We also thank Antoine Castet, Gwen-Jiro Clochard, Kai Gehring, Joseph Gomez, Eliot Green, Klaus Grundler, Jean-Marie Baland, Jeremy Lucchetti, Eoin McGuirk, Hannes Mueller, Alexandra Scacco, Mathias Thoenig, Ragnar Torvik, Anthony Venables, Pedro Vicente, Alexander Yarkin and participants at many seminars and conferences for useful discussions and comments. Mariama Sow provided outstanding research assistance. The project leading to this research has received funding from Excellence Initiative of Aix-Marseille University - A*MIDEX, a French "Investissements d'Avenir" program and from the French National Research Agency Grant ANR-17-EURE-0020, and by the Excellence Initiative of Aix-Marseille University - A*MIDEX. Mathieu Couttenier acknowledges financial support from the IDEXLYON from the University of Lyon (French National Research Agency, "Programme Investissements d'Avenir" ANR-16-IDEX-0005) and the Swiss National Foundation (grant Ambizione-PZ00P1 161451). Victoire Girard acknowledges funding by the Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722) and POR Norte. A previous version of this paper circulated under the title "Natural resources and the salience of ethnic identities".

†Aix-Marseille University, CNRS, EHESS, Centrale Marseille, AMSE, France, and CEPR. E-mail: nicolas.berman@univ-amu.fr.

‡University of Lyon, ENS Lyon, GATE Lyon/St-Etienne and CEPR. E-mail: mathieu.couttenier@ens-lyon.fr

§NOVAFRICA, Nova School of Business and Economics, Universidade NOVA de Lisboa, and LEO, Univ. Orleans. E-mail: victoire.girard@novasbe.pt.

1 Introduction

Social identities are a first-order determinant of a large group of social, economic, institutional and political outcomes (Alesina et al., 1999; Alesina and La Ferrara, 2005). In Africa, the high levels of social fragmentation resulting from identification to ethnic groups – rather than nations – has long been pointed as a primary cause of the "growth tragedy": identity fragmentation appears to dampen trust, cooperation, public goods provision and government performance (Easterly and Levine, 1997; Habyarimana et al., 2007, 2009). However, we are still a long way from understanding to which extent these groups' identities are fixed, and, if not, what makes them evolve (Anderson, 2006; Akerlof and Kranton, 2010).

This paper studies a potential driver of individual identification to ethnic groups in Sub-Saharan Africa (SSA): the exploitation of mineral resources. Mineral resources exploitation offers a unique setting to investigate the mechanisms underlying variations in ethnic identities, for three main reasons. First, minerals are point-source commodities, which are attached to specific locations. In SSA, these locations belong to historical, pre-colonial ethnic homelands, entities which still matter today and have been shown to influence current economic and political outcomes (Michalopoulos and Papaionnaou, 2020), despite colonization and subsequent nation-building policies. The exploitation of such resources is therefore intrinsically connected to the ethnicity question. Second, the perceived ownership of such resources is a recurrent subject of disputes between individuals who may claim the rights to ownership of natural resources – and therefore their potential rents (Collier, 2017): nationals from the country hosting the resource, and those who feel ownership of the land above the resource (the locals, or the members of the resource rich ethnic group). These disputes may escalate up to the emergence of secessionist parties or to secession attempts, as in Aceh, Biafra or Katanga (Aspinall, 2007; Gehring and Schneider, 2020). Finally, given the debate on who should benefit from resources rents, minerals' extraction is potentially associated with positive economic spillovers as well as with various grievances.[1] By studying how mining affects a variety of individual outcomes – wealth, grievances, or actual and perceived ethnic inequality – we can discuss *how* mineral resources affect ethnic identification, and therefore provide results which may inform us more broadly about the drivers of ethnic – or local – identification beyond the resources case.

To study the link between ethnic identities and natural resources, we combine three main sources of data covering 25 SSA countries between 2005 and 2015. First, we exploit geo-localized individual data on the salience of ethnic (*versus* national) identities from four survey waves of the Afrobarometer. Second, we match the information on the ethnic group of each Afrobarometer respondent with a list of historical ethnic groups, which gives us the geographical boundaries of the historical ethnic homelands. Last, we make use of time-varying information on the location and exploitation of mineral resources to compute the yearly number of active large-scale mines in the ethnic homeland of each Afrobarometer respondent – this number of mines is our baseline proxy of resources exploitation in a homeland. Equipped with these data, we estimate how respondents from different ethnic backgrounds respond to changes in the relative intensity of natural resources exploitation in their historical ethnic homelands.

Our baseline identification strategy exploits three dimensions of variation in exposure to min-

[1] For a literature review on the effects of natural resources on local development, see for instance, van der Ploeg (2011); Aragón et al. (2015); Cust and Poelhekke (2015); Venables (2016)

eral exploitation: time, geographic proximity and identity proximity. Consider two individuals living at a given point in time but belonging to different ethnic groups. Based on our combined data on the contours of each ethnic homeland and on time variations in mineral exploitation, we identify how changes in relative mining activity, over time and across the ethnic groups of these two individuals, affect their identification to their ethnic groups. The structure of the data allows us to partial out any local time-varying shocks affecting both individuals, and to account for any time-invariant differences in the level of identification to a given ethnicity in a given country, such as a group size or its historical political dominance.

We first document that ethnic identification becomes stronger when mineral resources exploitation intensifies in the historical homeland of the respondent's ethnic group. Interestingly, the effect of resources is significant both when exploitation takes place in the respondent's country of residence, and when it takes place abroad; this resonates with the idea that ethnic borders continue to matter in a continent where national borders largely ignore them (the "scramble for Africa", e.g. Michalopoulos and Papaionnaou, 2020). We also find that the effect of mineral resources on ethnic identification is persistent (and stronger two to three years after the exploitation starts), and tends to be magnified during electoral periods. The baseline estimates are robust to a large battery of checks, including various definitions of ethnic homelands, or using alternative measures of mineral resources exploitation, namely changes in the value of mining production generated by variations in the world prices of minerals. We also reproduce our main analysis using ethnic homeland specific variations in weather; we do not find any effect of such weather shocks, which suggests that our main results are driven by specificities of mineral resources extraction, rather than by general income fluctuations. Finally, we find that the effect of mining on the salience of ethnic identities holds regardless of whether the individuals live in their historical homeland. Overall, this first part of the paper shows that mineral exploitation causes more fragmented identities, by increasing the divergence in the importance of the ethnic or national identities reported by the members of the different groups present in each country.

The second part of the paper focuses on the channel underlying these findings. At the core of our discussion are the ethnic-group specific feelings of "relative deprivation" and grievances. New mining activities may raise aspirations related to economic development, in particular among members of the ethnic groups hosting the resources (feeling a sense of entitlement over the rent from resources extracted in their group's homeland, Collier, 2017). However, under incomplete information, these aspirations may be unrelated to the reality of resources' extraction, resulting in feelings of deprivation. We build upon specific case studies and recent literature linking economic development to aspirations and identities (notably Aspinall, 2007; Binzel and Carvalho, 2017; Cheeseman and Larmer, 2015; Must, 2018; Olzak, 2011). These deprivation feelings may heighten group identification, either in a decentralized manner (individuals valuing a social identity as a way to cope with unfulfilled economic aspirations, as in Binzel and Carvalho, 2017), or following an opportunistic manipulation by "ethnic political entrepreneurs" (who use ethnic grievances to divert attention from their own rent extraction, or gain political leverage, as in Aspinall, 2007; Cheeseman and Larmer, 2015). The main competing explanation would be that natural resources make ethnic identities more salient by increasing the economic or political "payoffs" associated with belonging to a particular group. This channel would be observed if minerals had strong positive local economic effects, that are felt disproportionately by historical local groups; or if

they increased the political power and representation of these groups; or political patronage along ethnic lines. The actual payoffs of belonging to a given a group would increase, which would make identities more salient. Combined with intra-ethnic spillovers (due to migration, trade, remittances, or solidarity), both the "payoff" and the "deprivation" channel could also explain why individuals residing outside their homeland also display stronger identity feelings. However, the payoff interpretation has a number of implications which differ from the deprivation channel, and which we can test empirically.

We provide a set of empirical exercises which globally support an interpretation rooted in the impact of natural resources on feelings of deprivation and ethnic grievances. We firstly show that in our sample, the effect of mineral resources on household wealth is limited, while the reports of economic deprivation and pessimism about economic conditions increase significantly with mineral exploitation. Interestingly, these effects are magnified for co-ethnics living outside their homeland, possibly because the discrepancy between the expectations and the reality of the economic spillovers from resources exploitation is magnified in this sub-group. Using information on mining companies' characteristics, we also find that domestic mines tend to magnify self reports of deprivation more than non domestic mines, although the effect on individual wealth is similar for both types of mines ownership. Second, we investigate how the local effects of mines vary with distance to mines, and across ethnic groups. We find that mines have limited local economic effects and, crucially for the deprivation channel, that there is no additional economic benefit for individuals belonging to the group whose historical homeland hosts mining activities. On the other hand, individuals belonging to that group report more deprivation when mining activity intensifies. Taken together, these results support the deprivation rather than the payoff channel.

Finally, we argue that, if indeed caused by deprivation feelings, the link between natural resources extraction and ethnic identities should be magnified in areas where feelings of ethnic inequality are pre-existing, e.g. relatively poor regions, politically excluded ethnic groups, or in areas with a history of violent conflicts (as appears in the case studies of Aspinall, 2007; Must, 2018). We bring these predictions to the data, making use of information contained in the Ethnic Power Relations (EPR) dataset on political exclusion, and of data on conflict and regional poverty. We show that the effect of mineral resources on ethnic identification is magnified both for groups without any ethnic political representation, and for powerless groups. We also find that a recent history of conflict in the ethnic homeland strengthens the mining-identity relationship.[2] Finally, we document that this relationship is magnified for members of relatively poor ethnic groups, as compared to the country average. These results corroborate the importance of the grievance channel in the mines-identity relationship.

Contribution. Our paper contributes to three main research strands. The first is the recent research emphasizing the endogenous nature of social identities. Since the seminal paper of Easterly and Levine (1997), numerous work have documented the economic consequences of fragmented identities (as reviewed in Alesina and La Ferrara, 2005). However, the strength of

[2]This results uses information on the *pre-existing* intensity of violent conflicts in the homeland, that is, conflicts that took place before the beginning of our study. The more general relationship between minerals, identities and conflicts is beyond the scope of this paper. Whether the exacerbation of ethnic identity triggers conflict, or the opposite, or whether these two variables are co-determined (both may be means of reaching the same objective for the leaders) remains an open question.

these identities feelings might not be fixed (Anderson, 2006; Akerlof and Kranton, 2010). We contribute to the recent line of research which shows that the existence of group identities, and the distance between them, are malleable.[3] Any individual has many social identities – based on nationality, gender, class, ethnicity, etc. In theory, which of these identities is the strongest can depend on how close one feels to other members of that identity relative to members of other identities, on the social salience of that identity, or on the payoffs one may expect from that identity (Shayo, 2009). Empirically, strong group identities have thus been associated with either major events affecting the distance between individuals and the salience of certain identity lines – such as an ethnic conflict, or a national sport victory (Rohner et al., 2013; Depetris-Chauvin et al., 2020) – or with the existence of positive group payoffs – such as an ethnic group being historically richer, or ethnic competition in political campaigns (Pengl et al., 2021; Eifert et al., 2010). Our paper contributes to this literature by highlighting how individuals react to a shock which is ethnic-group specific and time-varying. This setting allows us to draw attention on the importance of payoffs perceptions and grievances highlighted in case studies (Aspinall, 2007; Cheeseman and Larmer, 2015; Must, 2018), and to document how specific ethnic group characteristics are likely to reinforce the cycle between identity fragmentation, deprivation, and exclusion.

Our results also call for giving more attention to social identities and grievances in the large literature on the effects of natural resources' extraction (surveyed by van der Ploeg, 2011; Cust and Poelhekke, 2015; Venables, 2016). A body of work has documented the political economy side of the so-called "natural resource curse", and outlined the risks of conflicts about rent sharing across identity groups or regions.[4] Using data on mines in Africa similar to ours, Berman et al. (2017) find that increases in the world prices of minerals are associated to more violence locally, and that variations in the value of mines within specific ethnic homelands foster the activity of rebel groups related to these ethnic groups. These findings call for a better understanding of what motivates individuals living in resource-rich environments to engage in violence or support fighting groups. By documenting a link between natural resources, identity, and deprivation, our work can be viewed as an additional step in that direction. Recent works argue that already before exploitation starts, the existence of resources triggers an optimism unrelated to their economic reality, among both specialized growth forecasters and households (Cust and Mihalyi, 2017; Cust and Mensah, 2020). Here, we show that after resources exploitation starts, and even if resources may have positive economic spillovers (Aragón and Rud, 2013),[5] group-specific grievances and identity fragmentation arise. These results suggest a new political economy channel through which resources exploitation may hinder development.

Finally, our work also indirectly speaks to the literature on political borders in Africa, which

[3]On identity endogenization, see for example Klor and Shayo (2010); Rohner et al. (2013); Robinson (2014); Bisin et al. (2016); Ahlerup et al. (2017); Green (2018); Dehdari and Gehring (2022). On the importance of distances across identities on development processes, the most common measures of identities fragmentation, namely the polarization index and the fractionalization index, are weighted by distance (Esteban et al., 2012). The fractionalization index assumes by default a maximum value for distance (at 1). Given the difficulty to measure inter group distance, the same has often been done in studies using the polarization index. In the absence of any inter-group distance, these measures would take a value of zero, no mater how fragmented the society is.

[4]See for instance Monteiro and Ferraz (2010); Caselli and Michaels (2013); Morelli and Rohner (2015); Loayza and Rigolini (2016); Berman et al. (2017); Collier (2017); Fenske and Zurimendi (2017); Vogt (2017); Mamo and Bhattacharyya (2018); Christensen (2019); Lessmann and Steinkraus (2019); Girard et al. (2021).

[5]The relative deprivation we document, grounded in a discrepancy between expectations and reality, will be exacerbated if the forecasts of the resources economic spillovers are over optimistic.

shows that, partly because of the arbitrary way in which national borders were designed by colonial powers (the "scramble for Africa"), pre-colonial ethnic institutions persist and have still observable effects today (see the survey by Michalopoulos and Papaionnaou, 2020). We find that what occurs inside historical ethnic homelands matter, even for individuals living outside their historical homelands. We also find that natural resources in one's homeland have a significant effect on ethnic identification when the resources are located in a part of the homeland which is outside the respondent's country of residence. Put differently, our results provide new evidence supporting the prevalence of ethnic borders over administrative ones.[6]

The rest of the paper is organized as follows. Section 2 present the data and our baseline empirical strategy. Section 3 contains the results on how mineral resources activity shape ethnic identification. In Section 4, we discuss theory and evidence supporting the deprivation channel. The last section concludes.

2 Data and Empirical Strategy

To assess the impact of mineral resources on ethnic identification, we combine data on (i) the location and incidence of resources' production; (ii) the historical borders of ethnic groups' homelands; (iii) individuals' level of identification to their ethnic group. The appendix contains more details about the dataset construction.

2.1 Ethnic identification

We exploit rounds 3 to 6 of the Afrobarometer surveys.[7] The data are repeated cross-sections and cover a total sample of more than 100,000 respondents in 25 Sub-Saharan African countries over the period 2005-2015.[8] For each round, the Afrobarometer reports the exact centroid coordinate of respondents' town, village or neighborhood of residence (BenYishay et al., 2017), and provides detailed individual characteristics such as age, gender, education level, employment status, as well as ethnic group identification.

Our main outcome variable is the strength of individual feelings toward their ethnic group identity. It comes from the following question: *"Let us suppose that you had to choose between being a [National] and being a [R's Ethnic Group]. Which of the following best expresses your feelings?"* Our baseline variable of "ethnic identification" takes the inverse of the values suggested in the questionnaire, hence the highest values denote the strongest identity feeling: *5=I feel only*

[6]Our results align well with the literature stressing how ethnic networks and linkages (Fafchamps, 2000, 2003; Aker et al., 2014), ethnic leaders (Asiwaju, 1985; Aspinall, 2007; Michalopoulos and Papaionnaou, 2020), or sheer ethnic solidarity (Habyarimana et al., 2007, 2009; Lowes et al., 2015), all craft behaviors, and matters potentially as much or more than individuals' physical location.

[7]The question on ethnic identity that we exploit was asked consistently in these rounds, while earlier rounds gave respondents a list of several alternatives identities.

[8]The Afrobarometer coverage is not exhaustive and calls for caution in generalizing our results. However, we are confident that our results are informative about a significant share of the continent population as in 2015, 70% of the population of SSA was living in one of the countries included in our sample (population numbers source: the United Nations World Population Prospects 2019). The countries covered are: Benin, Botswana, Burkina Faso, Cameroon, Cote d'Ivoire, Ghana, Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mozambique, Namibia, Niger, Nigeria, Senegal, Sierra Leone, South Africa, Tanzania, Togo, Uganda, Zambia, and Zimbabwe (see appendix A1 for the year of survey of each country). Each survey round in each country has a nationally representative sample of either 1,200 or 2,400 respondents. Data are available at http://www.afrobarometer.org. The appendix reports the year of survey for each country.

(R's ethnic group), 4=I feel more (R's ethnic group) than national, 3=I feel equally national and (R's ethnic group), 2=I feel more national than (R's ethnic group), 1=I feel only national. Alternatively, we use a binary variable taking the values 1 if the respondent's answer is *I feel only* or *I feel more* or *I feel equally*. We also consider a more restrictive binary measure, which equals 1 only if the respondent feels more or only ethnic. In total, 53% of the respondents declare to feel at least equally strongly or more strongly about their ethnic identity than about their national identity.

The Afrobarometer surveys contain several other variables that we exploit to document the channels of transmission. In particular, we make use various proxies for wealth and economic deprivation, as well information about the respondents' perception of her economic conditions. Finally, we also exploit a question where respondents are asked whether they believe that their ethnic group is treated unfairly.[9] We describe these variables in section 4 and present their summary statistics in Appendix Table A.8.

2.2 Boundaries of ethnic homelands

We combine the Afrobarometer surveys with information on the boundaries of respondents ethnic homelands to assess individuals' reactions to shocks occurring in these homelands. Our baseline estimations exploit the digital maps of historical ethnic homelands of Nunn (2008), based on the compilation by Murdock (1959) of the ethnographic information available in the late nineteenth century. We match the self-reported ethnic group of respondents to a Murdock homeland for 89% of the respondents.[10] This results in a total of 354 ethnic homelands located in 40 countries.[11] These homelands are represented in Figure 1, which also displays the average level of ethnic identification by ethnic homeland, computed from Afrobarometer data. More precisely, the map shows, for each ethnic group, the share of members of that group who value their ethnic identity equally or more than their national identity: there is a substantial variation both across and within countries in the salience ethnic identification.

In our baseline estimations, we match each ethnic group of the Afrobarometer surveys with a single, main group from Murdock (1959). This matching procedure is the most direct, and also used by Nunn (2008) and Nunn and Wantchekon (2011). It is however restrictive: Afrobarometer and Murdock groups might not share a unique correspondence. An Afrobarometer group, for instance, might appear as several sub-groups in Murdock.[12] This explains why multiple blank areas appear in Figure 1, even in countries covered by the Afrobarometer. To take into account this limitation, we also match Afrobarometer and the Murdock Atlas homelands using an alternative matching rule allowing for multiple correspondence between the two datasets. We make use of an aggregated definition of groups present in the Murdock Atlas, namely the "cultural" groups.

[9] Another relevant variable relates to interpersonal trust, particularly to trust on co-ethnics versus other ethnic groups. Unfortunately, this variable only in one of the four Afrobarometer rounds that we consider, which makes it unusable in our context.

[10] We attribute to the respondent an ethnic group based on the following question: *"What is your ethnic community, cultural group or tribe?"* Only 0.5% of the respondents do not answer the question and 1% of them answer a national identity only.

[11] This number is larger than the 25 countries included in the Afrobarometer, as the homelands of the Afrobarometer's respondents often span over countries that are not necessarily in the Afrobarometer.

[12] Indeed, although the Afrobarometer questionnaire on ethnic groups is supposed to use the same classification as Murdock (1959), the overlap is far from perfect. The most extreme case may be in Ivory Coast where the Afrobarometer only records 5 different ethnic groups while the Murdock Atlas contains 35 distinct groups.

The Murdock Atlas attributes each of its 835 ethnic homelands to 104 distinct Cultural groups.[13] In this approach using Cultural groups, a single ethnic group of the Afrobarometer surveys can be matched with several Murdock homelands, ensuring a more exhaustive coverage of the Afrobarometer countries. Last, we resort to an alternative matching procedure based on the recently released LEDA: Linking Ethnic Data from Africa (Muller-Crepon et al., 2020). While our baseline matching resorts to historical information on groups evolution and ancestry (like Nunn, 2008), the LEDA relies on groups proximity after linking these groups to the Ethnologue language tree. The outcome of both the Cultural group matching and the LEDA matching appear in shown in Figures A.3 and A.4 in the appendix A2, where the matching procedures are described in more details.

Though the Murdock data is widely used by the literature (for example in Gennaioli and Rainer, 2007; Nunn, 2008; Michalopoulos and Papaioannou, 2013; Alsan, 2015), the fact that homelands are a snapshot from historical maps comes both as an advantage and as an inconvenient. On the negative side, settlement patterns may be outdated. Also, as homelands maps were built based on existing anthropological work, the level of accuracy and resolution varies for different regions in the continent (coastal areas are typically more accurately described than remote inland areas, see e.g. Michalopoulos and Papaionnaou, 2020 for a discussion). On the positive side, this historical snapshot alleviates concerns of ethnic group location being endogenous to the evolution of ethnic identities salience that we aim at explaining.

As an alternative strategy, we also match ethnic groups of the Afrobarometer to homelands from the Ethnic Power Relations dataset (EPR). The EPR dataset being based on both contemporaneous and time-varying information, the borders of the EPR ethnic groups are more likely to be affected by contemporaneous shocks than the Murdock data. To mitigate this concern, we compute EPR related information based on the EPR dataset for *before* the start of our analysis, in the year 2004. We also note that the focus of the EPR data on politically relevant groups leads to omitting many ethnic groups – and sometimes entire countries – from the EPR sample.[14] On the other hand, boundaries of the EPR data are *de facto* more accurate. Moreover, we make use of the richness of the EPR records on the political status of each group. These data allow us to estimate how our results vary across ethnic groups with different degrees of access to political power.

2.3 Mineral resources

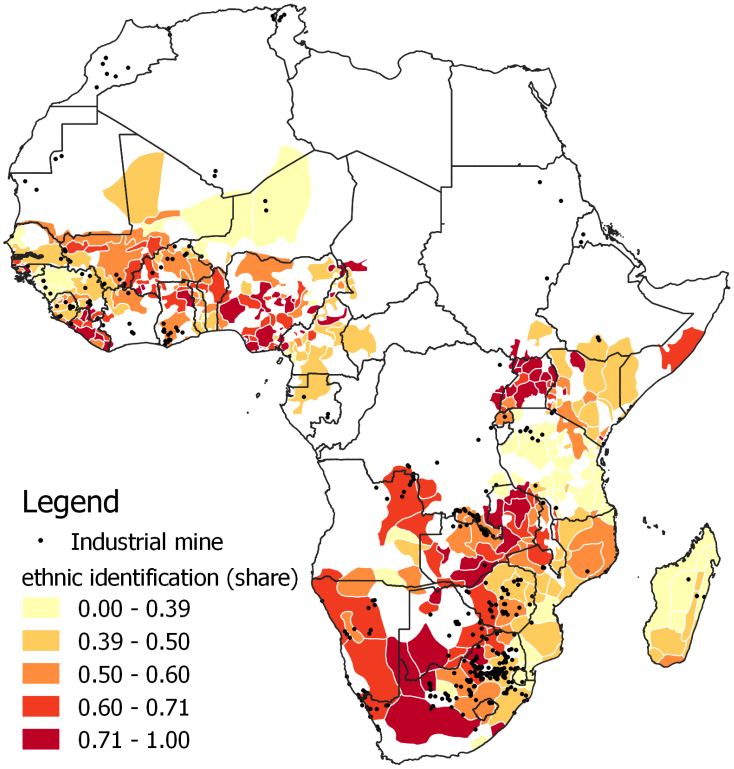
Data on natural resources come from S&P Global - SNL Metals and Mining. The dataset includes large-scale mines (industrial mines), usually owned and operated by multinationals or national firms. It covers 33 minerals, as well as information on the location of the mine, whether it is active, the volume of production, and the year production started.[15] We are therefore able to

[13]The "Lobi" cultural group territory will for example consist of the homelands of four ethnic groups, namely that of the Birifon, Dorosie, Kulango and Lobi.

[14]Vogt et al. (2015), p. 4, note that "An ethnic group is considered politically relevant if at least one political organization has claimed to represent its interests at the national level or if its members are subjected to state-led political discrimination (Cederman, Wimmer, and Min 2010, 99)."

[15]The following minerals are covered: U3O8, bauxite, chromite, chromium, coal, cobalt, copper, diamond, ferro chrome, gold, graphite, heavy mineral sand, ilmenite, iron, lanthanide, lead, lithium, manganese, molybdenum, nickel, niobium, phosphate, platinum, potash, rutile, silver, tantalum, tin, titanium, tungsten, vanadium, zinc, zircon. It omits artisanal and small-scale mines.

Figure 1: Ethnic homelands, mining and ethnic identification



Source: Authors' computations from the Afrobarometer surveys, Murdock and S&P data. See main text for data sources. The map shows the location of each industrial mine that has been active between the years 2005 and 2015 and the baseline matching of Afrobarometer and Murdock ethnic groups. For each ethnic group, the map shows the share of members of that group who value their ethnic identity equally or more than their national identity. Each share takes in account the answers of all the members of each ethnic group, independently of the homeland in which these members live.

compute the number of active mines in the ethnic homeland e of individual i at time t, which we use as our main explanatory variable. The location of mines is shown in Figure 1. Mining activity is clustered: on average, each ethnic homeland contains 2 mines, homelands with at least a mine contain on average 5 mines.

2.4 Descriptive statistics

Table 1 displays descriptive statistics on our final sample made of 115,105 respondents (the sample of our baseline estimates, column 1, Table 2). Most respondents live outside their ethnic homeland and in rural areas (54% and 63%, respectively). Respondents split evenly across genders, 61% of them have completed a primary education degree (or more), and 64% are active (either employed or looking for a job).

The respondents are divided across 296 ethnic groups in the final sample, with 20.9% of these ethnic groups hosting at least one mineral activity in their homeland over the 2005-2015 period. As resource rich ethnic groups are on average larger than resource poor ethnic groups, 34% of the respondents report coming from a group whose homeland hosts minerals.[16] The average number of mines substantially varies when we move from the most restrictive to the broader homeland definition: the average number of mines in our final sample is 2.1 with our baseline measure, and

[16]The complete sample – just after matching the self-reporting ethnic group of respondents and Murdock homelands – covers 354 groups. In this complete sample, 30.8% of the respondents report coming from an ethnic group rich in minerals.

goes up to 17.9 when we use Murdock's more aggregated "cultural groups" definition. Note that, as the mine variable is right-skewed, we check the robustness to the presence of outliers in our sensitivity analysis.

Table 1: Summary statistics for the main variables

	Mean	S.D.	1st Quartile	Median	3rd Quartile
Ethnic identity	2.35	1.19	1	3	3
Ethnic identity (dummy)[a]	0.53	0.5	0	1	1
Ethnic identity (dummy, alt.)[a]	0.13	0.34	0	0	1
# mines in homeland	2.06	5.79	0	0	1
# mines in homeland, LEDA homelands[b]	1.49	5.14	0	0	1
# mines in homeland, Cultural homelands	17.85	35.31	0	3	14
# mines in homeland, EPR homelands[b]	6.73	22.22	0	0	1
# mines in homeland, in country	1.36	4.93	0	0	0
# mines in homeland, abroad	0.71	3.24	0	0	0
Residence in homeland (dummy)	0.46	0.5	0	0	1
Residence in rural area (dummy)	0.63	0.48	0	1	1
Female (dummy)	0.5	0.5	0	0	1
Primary education or more (dummy)[c]	0.61	0.49	0	1	1
Active (employed or looking) (dummy)[c]	0.64	0.48	0	1	1
Age	36.69	14.49	29	33	45
Number of respondents = 115,105					

Source: Authors' computations from the Afrobarometer surveys, Murdock and S&P data. See main text for data sources. [a] Ethnic identity (dummy) equals 1 if the respondent's answer is *I feel only* or *I feel more* or *I feel equally* ethnic. Ethnic identity (dummy, alt.) equals 1 only if the respondent feels more or only ethnic. [b] The descriptive statistics for the LEDA matching come from the sample of 116,117 households who belong to groups identified in the LEDA dataset, the descriptive statistics for the EPR come from the sample of 84,841 households who belong to groups identified in the EPR dataset. [c] In our estimations, we control for all nine education categories and all four employment categories provided by the Afrobarometer surveys.

2.5 Identification strategy

The purpose of our empirical strategy is to estimate how natural resources extraction - in the form of minerals - taking place in the historical homeland an ethnic group affects the strength of identity feelings declared by the members of that ethnic group. Our source of identification relies on the comparison of two individuals living in the same region, in a given point of time, but belonging to two different ethnic groups.[17] Combining our data on the contours of each individuals' ethnic homelands and the time-varying information of mining activity location, we identify how changes in relative mining activity across each individuals' homelands, over time, affect their relative ethnic identification:

Formally, for an individual i belonging to ethnic group e, living in region r of a country c at time t (the specific month of each year), we estimate the following specification:

$$\text{ETHNIC IDENTIFICATION}_{i,e,r,t} = \alpha \times (\#\text{MINES})_{i,e,t} + \mathbf{C_i}'\beta + \mathbf{FE}_{e,c} + \mathbf{FE}_{r,t} + \varepsilon_{i,e,r,t} \quad (1)$$

The dependent variable represents the strength of ethnic identification of individual i (as a cate-

[17] This approach is fundamentally distinct to the paper maybe closest to us, Ahlerup et al. (2017). Ahlerup et al. (2017) document how local income, proxied by nighttime light and instrumented by mineral production, is associated with more identification to the nation. Our work differs in terms of research question (taking resources as a group-specific shock rather than considering their average effect), identification strategy (allowing resources to matter through other channels than a wealth effect), and sample (extended in time and space, allowing us to control for any time-invariant ethnic group characteristics, such as pre-existing differences in the level of identification across groups).

gorical or as a dummy variable), and our main explanatory variable is the number of active mineral resources exploitation in the ethnic homeland e of individual i at time t ($\#\text{MINES}_{i,e,t}$). We test the robustness of our results to using various definitions of ethnic homelands. We consider two different matching procedures between the Afrobarometer's ethnic groups and Murdock's homelands, as explained in section 2.2. We also show results using the alternative contours of homelands boundaries from EPR.

The vector $\mathbf{C_i}$ includes a large set of respondent characteristics: gender, age and its square, a set of dummies controlling for education levels and employment status, and residence in a rural area. We also control for a dummy coded one if individual i is living in her ethnic homeland at the time of the survey, and zero otherwise. We include ethnic homeland-country fixed effects ($\mathbf{FE}_{e,c}$), which capture any time-invariant feature of each ethnic group in each country, as well as inherent differences in the level of identification to a given ethnic group in each country (and other historical aspects like its historical political dominance). We also include region of residence $\times$ year $\times$ month fixed effects ($\mathbf{FE}_{c,r,t}$), where a region is defined at the first administrative level (ADMIN-1).[18] This allows us to account for local shocks such as climate variations or crop prices. Any local economic spillover of natural resources exploitation that would be specific to a location, without affecting deferentially the members of the different ethnic groups, is also captured by $\mathbf{FE}_{c,r,t}$. Given the inclusion of these fixed effects, α identifies the effect of mining activity in individual i's homeland on that individual's ethnic identification, relative to individuals of a different ethnic group living in the same region at a given point in time. The inclusion of two sets of high dimensional fixed effects leads us to use a linear probability model to estimate (1), to avoid the incidental parameters problem. Finally, standard errors are clustered at the ethnic group level, though we show that our baseline results are extremely similar when allowing for spatial correlation in the error term.

2.6 Identification issues

Saturating the model with a large array of fixed effects limits omitted variables concerns and eases interpretation, but the remaining variance of our variable of interest $\#\text{MINES}_{i,e,t}$ may become small, and affected by specific cases. We consider this issue in section A5 of the appendix, where we show how the residual variation in the number of mines is affected by our various dimensions of fixed effects. As shown in Table A.10, the raw variable has a standard deviation of 5.74. When purged from the contribution of our baseline set of individual controls $\mathbf{C}_i$, the standard deviation is barely affected (5.68). The variation drops by much more when ethnic homeland or ethnic homeland$\times$country fixed effects are included: the standard deviation goes down to around 1. This is expected, as these fixed effects capture all the variation in the homelands in which the number of mines does not change over our sample period - which is about 80% of the homeland$\times$countries. Interestingly, the standard deviation remains high (4.21) when the region of residence $\times$ time fixed effects are included. This is because individuals of different ethnic groups

[18]The Afrobarometer survey being a repeated cross section, the exact village or town neighborhood where the survey takes place moves across waves. Making use of the the exact centroid coordinate of respondents' town, village or neighborhood of residence to define the region of residence would result in more than 12 000 local $\times$ year $\times$ month fixed effects (median of 5 respondents by level of fixed effects). We thus instead use the region of residence defined at the regional (ADMIN-1) level. In our sensitivity analysis, we show that our baseline results are robust to using a more disaggregated definition of a region, or a definition of region that does not rely on political boundaries.

are present at this level, making it more likely that a different number of mines is exploited in their homeland. When all fixed effects are included, the standard deviation falls at 0.77. Hence, thinking about the effect of "one additional mine" – the metric we use when commenting our results – is not very far from a (conditional) s.d. increase in the number of mines. Note that we discuss in the robustness section 3.3 how sensitive our results are to using more or less restrictive sets of fixed effects.

Finally, a potential endogeneity concern is that observed changes in mining activity over time – opening and closing of mines – could be driven by ethnic groups' behaviors that correlate with the strength of ethnic identities. For instance, the political power of the ethnic group might both drive ethnic identification and changes in natural resource production. We will show that many of our results are difficult to reconcile with this view: in fact, the strongest effect is found for the least politically relevant groups. The number of active mines is also easy to observe for citizens. Still, in our robustness exercises we use an alternative measure of mining activity, which combines information on the volume of production at the beginning of the period with (exogenous) yearly variations in world prices of minerals.

3 Mineral resources activity and ethnic identification

3.1 Baseline estimates

Table 2 displays the baseline estimates. All columns include country × ethnic group fixed effects and region of residence × year fixed effects, where a region is defined at the Admin-1 level and time corresponds to month-year. We also control for a set of respondents' characteristics – gender, education, whether the respondent is in a rural or urban area and whether the respondent is located in her historical homeland. Table A.9 in the appendix reports the complete set of coefficients for these control variables. In line with the literature (Robinson, 2014), female, uneducated and rural respondents tend to identify more to their ethnic group.

We start by estimating the impact of the number of mines in the respondent's ethnic homeland on the categorical measure of ethnic identification (column 1). The measure ranges from 1 to 5, with higher levels denoting stronger identification of the respondent to her ethnic group. The coefficient on the number of mines is positive and significant: ethnic identities become more salient as the number of mines in the ethnic homeland increases. Quantitatively, an additional mine raises the level of ethnic identification by 0.018 (column 1), i.e. around 1.5% of the (unconditional) standard deviation of ethnic identification. Though it appears limited at first glance, the order of magnitude of the estimated effect of mining is comparable to that of other documented determinants of ethnic identities (Robinson, 2014). For instance, female respondents, inactive individuals, those living in a rural area, or residing in the historical homeland of the ethnic group respectively exhibit 0.051, 0.050, 0.032 and 0.044 higher levels of ethnic feelings (see appendix Table A.9, column 1), i.e. between 2.7 and 4.3% of the standard deviation of the ethnic

Table 2: Natural resources extraction and ethnic identification

	(1)	(2)	(3)	(4)	(5)
Dep. Var.	Salience of ethnic group identification				
Version	Categorical	Dummy	Dummy (alt.)	Categorical	Categorical
Sample	Full	Full	Full	Full	Outside homeland
# mines: ethnic homeland	0.018^a	0.006^a	0.004^a		0.016^a
	(0.003)	(0.002)	(0.001)		(0.003)
# mines: ethnic homeland, in country				0.017^a	
				(0.005)	
# mines: ethnic homeland, abroad				0.018^a	
				(0.005)	
Residence in homeland	0.044^a	0.014^a	0.012^a	0.044^a	
	(0.012)	(0.005)	(0.003)	(0.012)	
Fixed effects	Country×Ethnic group, Region×Time				
Individual controls			Yes		
Observations	115105	115105	115105	115105	61790
R^2	0.185	0.192	0.110	0.185	0.180
Sample mean / s.d. dep. var.	2.35/1.19	0.53/0.50	0.13/0.34	2.35/1.19	2.31/1.18

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In column (5), we restrict the sample to individual living outside their ethnic homeland. The dependent variable is the level of ethnic identification with respect to state identification. In columns (1), (4) and (5), the variable ranges from categories 1 to 5. In column (2) we use as dependent variable a dummy taking the value 1 if the individual identifies at least as much to the ethnic group than the country. In column (3) we use an alternative definition of the binary dependent variable, taking the value 1 if the individual identifies strictly more to the ethnic group than the country. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. There are 1824 Admin-1 region×year×month fixed effects and 401 country×ethnic group fixed effects. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the sample of the corresponding column.

identification variable.[19,20] The effect is qualitatively similar when using the binary measure of identification, that takes the value one for all respondents who report feeling "only ethnic", "more ethnic than national", or "equally ethnic and national" (column 2), or alternatively only the first two categories (column 3).

It is well known that the historical homelands of many ethnic African ethnic groups have been partitioned by the national borders drawn by colonial powers: in our sample 167 of the 296 ethnic homelands span over more than one country. In column (4), we allow the effect of natural resources production to differs depending on whether it takes place in a part of the respondent's homeland which lies inside or outside the respondent's country. We do not detect any significant difference. Note that these results should not be seen as a comparison within ethnic groups: less

[19]We do not know since when an individual resides outside of her historical homeland, whether it is due to the evolution in time of the boundaries of the homeland or to a decision to move away taken by the individual or her ancestors. If an individual decision, next to usual motives like joining an urban center or a job opportunities, identity feelings might drive self selection into migration. If this were the case, we would expect individuals with least attachment to their ethnic identity to be the ones most likely to move away from their ethnic group homeland. This could contribute to explain the positive relationship that we observe between residence in one's group historical homeland and the strength of ethnic identity. In this scenario, the coefficient we estimate for residence in the homeland would be upward biased. The fact that the magnitude of the effect of active mines is about half of that of residence in homeland would be an even stronger signal of the importance of our channel on ethnic group identification.

[20]Quantitatively, our baseline effect is also comparable to the one found by Depetris-Chauvin et al. (2020), who show that ethnic identification decreases in the days following sport victories of the national team. In their specification closest to our baseline (Table A.11, col. 2) a national team's victory decreases the level of ethnic identification by around 3.9% of the standard deviation.

than 3% of the respondents in our sample have an ethnic homeland partitioned across national borders and hosting natural resources on both sides of the border. Hence, the coefficients shown in column (4) are identified across rather than within groups, which prevents us from comparing their relative size. The fact that mines outside the respondent's country are found to significantly affect ethnic identification is however informative as it echoes the literature on political borders in Africa, according to which ethnic partitions are still relevant entities today. This result also aligns with the general story we uncover in the mechanisms section – that natural resources foster ethnic identification by triggering feelings of exclusion and economic deprivation for relatively powerless groups: indeed, those feelings may arise regardless of whether the homeland of the individual is located within or outside the country. These feelings may travel through political borders as ethnic networks and linkages, ethnic leaders, or sheer ethnic solidarity craft behaviors and may at times be more relevant than (often artificial) national borders.[21]

Our estimations, so far, do not make any distinction between mines operating within the region of residence of the individuals or outside that region. These two cases are however quite different, if only because local mines may directly impact local economic conditions. In column (5), we restrict the sample to respondents who live outside of their ethnic homeland. The magnitude of our coefficient of interest barely changes.[22]

Our interpretation of these estimates is that ethnic identity is likely to become increasingly salient for individuals belonging to "resource rich" ethnic groups – whose homelands host natural resources –, as opposed to national identity becoming more salient for resouce-poor individuals. We focus the interpretation on resource-rich individuals for two main reasons. Firstly, when minerals are located abroad, minerals should not impact identification of the individuals who are co-nationals of the members of the resource-rich groups, but without having any resources themselves. The significant effect of resources abroad which appears in column (4) of Table 2 is thus likely driven by an increase in ethnic identification of resource-rich individuals. Second, though we do not fully rule out that members of resource-poor group may react to the presence of minerals within their country,[23] intuitively, we expect members of resource-rich groups to react to resources exploitation at least as much as the members of other groups.

Overall, our estimates correspond to an increase in identities fragmentation, as we document

[21] Co-ethnics living on both sides of a country's border may share strong economic linkages given the importance of ethnic networks in the credit market, for trust, trade or migration (Fafchamps, 2000, 2003; Aker et al., 2014). Co-ethnics may also refer to common ethnic leaders and institutions. As an illustration, Asiwaju (1985) cites a Ketu king claiming during colonial times "we regard the boundary [between Benin-Dahomey and Nigeria] separating the English and the French, not the Yoruba." These ethnic institutions and leaders have endured the colonial and post-independence era and are, still today, the relevant unit of projection for many individuals (Michalopoulos and Papaionnaou, 2020). Furthermore, some leaders may strategically manipulate ethnic identities (Aspinall, 2007, acting as "ethnic political entrepreneurs"). Finally, co-ethnics living in different countries might share their deprivation feelings through the pure empathy or solidarity effect documented in the experimental literature, which is independent from any personal or economic ties (Habyarimana et al., 2007, 2009; Lowes et al., 2015).

[22] We also allow the effect of natural resources to vary with individual characteristics such as gender, education or type of residence; the estimates do not display significant differences along these dimensions.

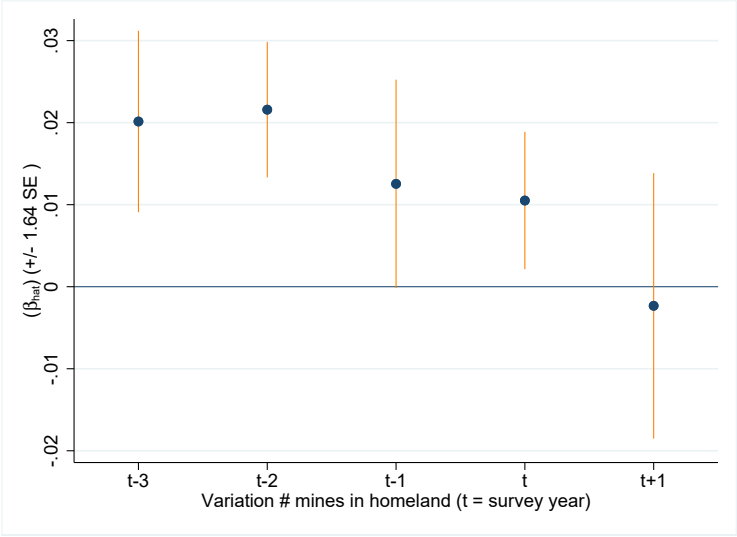
[23] The reaction of members of resource-poor groups could magnify our estimates if, as these individuals cannot claim the natural resources wealth through ethnic linkages, they use their national identity to justify that resources rents must be shared across all citizens. Alternatively, the reaction of resource-poor individuals might attenuate our estimates if members of all groups witness an increase in their ethnic identity feelings when exploitation intensifies, and the only reason we still identify a divergence is because the increase is strongest for members of the resource rich groups. A particular case that could lead to such a reaction is the one in which two groups contest an ethnic border with a mine; we could then expect that their members feel more ethnic given this contest. Such a contest would play against finding a significant effect, as it would lead to attenuation bias.

an increase in identities distance between the resource rich group members and other individuals. To be more precise, identity fragmentation is in theory a combination of the existence of identity boundaries between two or more groups, and the pair-wise distance between the members of these different identities (where various definitions of the identity groups are possible, such as ethnicity, religion or nation, Esteban and Ray, 2011). We here document a widening of the distance between the identities of the resource-rich and the resource-poor groups, a concerning finding in terms of comparative development.[24]

3.2 Persistence

Timing of the effects. The results of Table 2 show how individuals react to variations to the stock of active mines in their homeland, but are silent on the moment when the mining activity affects individual ethnic identification. Indeed, given that the rounds of the Afrobarometer occur several years apart, the changes in mineral exploitation could have occurred at the time of the survey or several years before. In other words, our baseline estimates could reflect either a short-lived "news shock" effect, or a more persistent one. To test when the impact of mining activity on ethnic identification starts taking place, we turn to a specification considering the time since the change in the number of mines.

Figure 2: Timing of the effect of natural resource extraction on ethnic identification



This figure reports the results of an estimation akin to equation (1), except that instead of including the number of active mines at the time of survey, we include variables representing variations in the number of active mines in the year of the survey as well as in the previous and subsequent years. We estimate the following equation: $\text{ETHNIC IDENTIFICATION}_{i,e,r,t} = \sum_{l=-1}^{3} \alpha^{t-l} \times \Delta\text{MINES}_{i,e,(t-l,t-l-1)} + \mathbf{C_i}'\beta + \mathbf{FE}_{e,c} + \mathbf{FE}_{c,r,t} + \varepsilon_{i,e,r,t}$, where $\Delta\text{MINES}_{i,e,(t-l,t-l-1)}$ is the yearly change in the number of active mines in the homeland e of individual i interviewed during year t, considering changes that took place up to three years before the year of the interview and a year after. The figure depicts the coefficients ($\hat{\alpha}$) and 90% confidence intervals.

Figure 2 shows the impact of changes in mining activity taking place from three years before

[24]To be more precise, our results may be documenting different situations of concern. First, we may be documenting an increase in distance between two groups in the country. For example, an increase in the distance between the resource-rich group – whose ethnic identity becomes more salient – and all other groups – whose identities remain unchanged, or more leaning toward their nation. These dynamics will lead to a more polarized society, a potentially problematic situation, in particular for the resource-rich group members if the rest of the nation turns against them (e.g. Morelli and Rohner, 2015). Alternatively, we may be documenting an increase in the distance between all ethnic groups in the country, for example, between the Mossi, the Lobi, the Fula, etc. leading to a more fractionalized society.

the survey up to a year after the survey. The estimation strategy is the same as in column (1) of Table 2, but we replace the variable representing the stock of the number of active mines at the time of the survey with separate variables representing the contemporaneous, past, and future changes in the yearly number of active mines.[25] The figure shows estimated coefficients as well as 90% confidence intervals. We find that the effect of changes in mining activity on identification to one's ethnic groups is quite persistent: in fact, the effect gets reinforced over time.

Before the start of mining operations, a phase of one or two years is required for investment (Benshaul-Tolonen, 2019). Locations where a mine started producing the year of the survey could already witness mine related activity (and most likely media coverage) the year before the start of the production. Yet, as shown in Figure 2 we detect no significant effect of changes in mining activity occurring a year following the survey. Hence, production appears to matter *per se*, not the prospects of production (Arezki et al., 2017).

Electoral periods. We pursue an alternative strategy to study persistence: we match our data with information on the timing of presidential elections and estimate whether the proximity to elections affects our estimates.[26] The underlying idea is that, if the effect is stronger during electoral periods, natural resources might have persistent impacts on ethnic identities through voting patterns, even if their impact is otherwise temporary. In the spirit of Eifert et al. (2010), we interact our main variable with the (log-)distance to the closest election (column 1 of Table A.11). We find that the impact of changes in the number of mines on ethnic identity is larger in proximity to elections. Alternatively, we interact our main variable with dummies equal to one if an election takes place in the 6 months, 6 to 12 months and more than 12 months, around the survey. The effect is stronger in the six months around elections and significantly different from the other periods (column 2). Last, the effect of the distance to the election does not depend on whether the closest election took place before or after the interview (column 3).[27]

3.3 Sensitivity analysis and additional results

In this subsection we test the robustness of the baseline estimates of Table 2 to a variety of sensitivity checks.

Alternative level of fixed effects. Our baseline model includes a restrictive set of fixed effects, at the country × ethnic group as well at at the region of residence × year × month level. A natural concern is that the variation in the number of mines left after including all these fixed effects is limited, which complicates the generalization of the results. Appendix A5 investigates this issue. We already discussed previously Table A.10, which shows that the residual variation in

[25] This approach would be equivalent to using leads and lags of the stock if we had yearly data with a large variance in the explanatory variable. However, the Afrobarometer provides a repeated cross-section with gaps of 3 or 4 years between surveys rather than a yearly panel. Moreover, about 50% of the groups with changes in the number of mines witness only one change over the survey rounds.

[26] To construct our measure of proximity to elections, we make use of the recently released Database of Political Institutions 2020 (DPI 2020, available at https://www.iadb.org/en/research-and-data/dpi2020). Similar to Eifert et al. (2010), we focus on presidential elections, which dates correspond to the dates of executive elections in the DPI 2020, complemented with the date of legislative elections for the three countries from our sample where the Assembly elects the president (namely Botswana, Lesotho, and South Africa).

[27] Appendix A6 shows similar results when we change the definition of the dependent variable for our baseline dummy of identification to ethnic group identities.

Table 3: Natural resources extraction and ethnic identification during electoral periods

	(1)	(2)	(3)
Dep. Var.	Ethnic group identification		
Version	Categorical		
# mines: ethnic homeland, in country	0.081^{b}		0.230^{a}
	(0.035)		(0.080)
# mines: ethnic homeland × Dist. elec.	-0.009^{c}		
	(0.005)		
# mines: ethnic homeland × Elec. [0-6months]		0.044^{b}	
		(0.018)	
# mines: ethnic homeland × Elec. [6-12months]		0.023	
		(0.015)	
# mines: ethnic homeland × Elec. [12+ months]		0.026^{c}	
		(0.014)	
# mines: ethnic homeland × Dist. elec. (after)			-0.014^{b}
			(0.006)
# mines: ethnic homeland × Dist. elec. (before)			-0.016^{a}
			(0.006)
# mines: ethnic homeland, abroad	0.018^{a}	0.019^{a}	0.018^{a}
	(0.005)	(0.005)	(0.005)
Fixed effects	Country×Ethnic group, Region×Time		
Individual controls	Yes		
Observations	109122		
R^2	0.187		
Sample mean / sd dep. var.	2.36/1.19		

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. The dependent variable is the level of ethnic identification with respect to state identification in all columns, ranging from categories 1 to 5. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. There are 1767 Admin-1 region×year×month fixed effects and 387 country×ethnic group fixed effects. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

the number of mines remains reasonable (a standard deviation of .77) after purging the variable from all our control variables and fixed effects. Figure A.5 shows how the coefficient of interest (Table 2, column 1) varies when including more or less restrictive sets of fixed effects. In all estimations, we include ethnic group × country fixed effects to ensure that the coefficient is identified across ethnic groups (hence country fixed effects are de facto included). Two conclusions emerge. First, it is important to control for time-varying local shocks. Controlling only for year dummies leads to insignificant results, and a coefficient close to zero. It is likely that both natural resource exploitation and ethnic feelings have a strong time-varying country-specific component.[28] Second, as long as we control for time-varying local shocks, the definitions of "time" (year or year-month) and "local" (country, admin-1 region, admin-2 region, or apolitical subdivisions, i.e. cells of 1×1 or 0.5×0.5 degree of latitude and longitude) do not affect much the precision of the

[28]Though smaller in magnitude and less precisely estimated, we do find significant effects when controlling for year × month fixed effects only. Yet, year × month fixed effects are likely to capture part of country × year shocks, as the month of survey is partly country-specific.

estimates, despite the drastic changes in the number of fixed effects (from 92 when including only country × year fixed effects to 1820 with country × region × year × month and 5852 with cells of 0.5×0.5 fixed effects). This result is consistent with the absence of strong local effects of mining as discussed in Section 4.2, and is reassuring: considering a wider set of variations in the number of mines produces results similar to our baseline, both qualitatively and quantitatively.

Endogeneity of mining activity. As mentioned earlier, variations in minerals' extraction in a specific homeland could be a function of the ethnic group's political power, which in turns might correlate with ethnic identification. This could bias upward our estimates. We consider this concern as quite unlikely given the findings discussed in Section 4 and in particular in Table 6: our estimates appear to be mostly driven by politically powerless groups. Still, to further ensure that our results do not reflect such potential endogeneity bias, we use an alternative measure of mineral resources (appendix, section A7). For each mineral produced in the ethnic homeland of a respondent, we compute the average pre-sample production (2000-2004, or 2004) evaluated at the yearly world prices of the mineral. Then, we sum across minerals to get the production value in the ethnic homeland e at year t: $Y_{e,t} = \sum_{k=1}^{N} Q_{e,k,t_0} \times P_{k,t}$, where Q_{e,k,t_0} is the production of mineral k in the ethnic homeland e during the reference period t_0 (here the average pre-sample production), and $P_{k,t}$ is the real world price of the mineral k during year t. The data on (real) world prices come from the World Bank Commodity Dataset. The inclusion of ethnic group × country fixed effects absorbs differences in the levels of production at the beginning of the period, hence the identification relies on yearly variations in world commodity prices which are arguably exogenous to local ethnic identification, especially given that the countries in our sample are typically small producers at the world level.[29] A concern with this measure is that individuals have arguably less information about changes in the world value of minerals than about the activity status of a mine.[30] Overall, our results are however globally robust to this alternative measure of mining activity (appendix Table A.12). We find positive coefficients, with p-values around 0.11-0.15[31], depending on the columns and definitions. Quantitatively, the effects are close to our baseline estimates: for instance, a standard deviation increase in production in column (1), triggers an increase in ethnic identification equivalent to 1.3% of the standard deviation of the dependent variable (compared to 1.5% for an additional mine in Table 2, column 1).

Weather shocks in ethnic homelands. Are our results specific to mining, or do they reflect a more general effect of income shocks? In the appendix Table A.13, we add to our baseline estimations measures of homeland-specific weather shocks. We include, sequentially or jointly, measures of aggregate rainfall and average temperature (columns 1 to 3) and measures of anomalies for rainfall and temperature (columns 4 to 6).[32] The estimates are statistically insignificant.

[29] See Berman et al. (2017) for a discussion on this question using similar data.

[30] Even today, local populations may be unaware of some important economic shocks. More that half of the respondents in Armand et al. (2020) have limited knowledge about the recent major discoveries of gas offshore of their region. A possible solution could have been to instrument the number of mines with variations in world prices; but this strategy cannot be implemented, because we do not observe the set of minerals that could have been produced in places before the mines open.

[31] The coefficients in columns (1) and (3) become significant at the 1% level when we correct for spatial correlation in the error term instead of our baseline specification (clustering standard errors at the level of the ethnic homeland).

[32] Temperature and rainfall information are aggregated at the Murdock homeland level from the Climatic Research Unit gridded Time Series that provide climate dataset at a resolution of 0.5×0.5 latitude and longitude

These results are useful for two reasons. First as a robustness: these tests further ensure that our baseline results are not caused by omitted factors that might correlate with mining activity. Second and more importantly, the fact that variations in climatic conditions do not appear to affect the salience of ethnic identification suggests that our baseline results are indeed driven by specificities of minerals' extraction, rather than economic shocks in general.[33] We come back to the question of economic shocks in the next section.

Placebo. To further rule out the possibility of false positive in the main results, we run a placebo test as follows: for each year in the sample, we randomly assign the number of mines across ethnic homelands, and re-estimate specification (1) of Table 2 with this random variable. We repeat this Monte Carlo procedure in 1,000 draws. Figure A.6 of the appendix plots the sampling distribution of the obtained coefficient, compared to our baseline coefficient drawn as a red vertical axis (0.018). Reassuringly, the Monte Carlo coefficients are distributed far from their baseline estimates.

Sensitivity to specific groups, countries, time periods and local effects. In the Appendix, we show that our results are not driven by specific subsets of observations. We first drop countries and survey waves one by one (Figure A.7 and Table A.15, respectively). The results are remarkably stable. Figure A.8 displays the estimated coefficients for each ethnic group separately: 61% of the coefficients are positive (70% if we only consider the set of coefficients statistically significant at 10%). We come back to ethnic-group heterogeneity in section 4.3. Second, we investigate the robustness of our baseline results to the exclusion of outliers (Table A.16).[34] Last, given the spatial dimension of the data, we allow the error term to be spatially correlated, and auto-correlated. The standard errors of our variable of interest remains stable (Table A.17); for instance, the standard error of column (1), Table 2 remains in the interval 0.004-0.005 (compared to 0.004 in the baseline) depending on the spatial radius chosen (up to 1000 km).[35]

Alternative matching of ethnic-groups' homelands. As our results rely on the match between the historical ethnic homelands recorded in the Murdock Atlas and today Afrobarometer groups, we replicate our baseline estimates for the three alternative definitions of ethnic homelands discussed in section 2.2. First, the broader cultural groups recorded in the Murdock Atlas,

(Ian Harris and Lister, 2020). Data from https://crudata.uea.ac.uk/cru/data/hrg/. Introducing the square of these weather variables does not change the conclusion of these estimations.

[33] Weather shocks are significant economic shocks (Barrios et al., 2010; Dell et al., 2012). In our sample, and in line with existing literature, we find that rainfall shocks in the historical homeland of a respondent improves the material conditions of that respondent (measured through the wealth index, Table A.14).

[34] In columns (1) and (2), we consider the log and the inverse hyperbolic sine transformation of the number of mines in the ethnic homeland of the respondent. From columns (3) to (5) we exclude observations that are 3, 2 and 1 standard deviation away from the residual mean. Then, we exclude observations with high leverage, i.e. when the individual leverage is superior to 2k=N (with k being the number of predictors and N the number of observations, column 6). Finally, we exclude observations that shift the estimate at least to one standard error (dfbeta, column 7), and observations that shift the estimate at least to 4/N (Cook's distance, column 8).

[35] We use the Stata routine developed by Colella et al. (2020). More precisely, we apply a spatial HAC correction to our standard errors, allowing for both cross-sectional spatial correlation and location-specific serial correlation (Conley, 1999). Imposing no constraint on the temporal decay for the Newey-West/Bartlett kernel that weights serial correlation across time periods, the horizon at which serial correlation is assumed to vanish can be in finite (i.e., 100,000 years). In the spatial dimension, we estimates the standard errors using alternatively thresholds at 250, 500, 750 or 1000 km.

which ensure an exhaustive geographical coverage. Second, the LEDA matching from Afrobarometer ethnic groups to Murdock ethnic homelands, which relies on the Ethnologue language tree. Third, we turn to a matching of the Afrobarometer ethnic groups to the contemporaneous ethnic homelands as recorded in the EPR. With few exceptions, our results are qualitatively unchanged, and quantitatively similar for the LEDA matching, but coefficients are smaller for two alternative definitions (Tables A.18 and A.19). This is because the size of the homelands – and therefore the variation in the number of mines contained in them – varies across definitions. Using our baseline specification with the categorical measure of ethnic identification (Table A.18), we find that an increase of one additional mine raises ethnic identification by 0.005-0.007 when using the cultural group definition; 0.017-0.018 when using LEDA; and 0.007-0.009 when using EPR. But considering a standard deviation increase in the number of mines (after purging the variable from all fixed effects and controls) leads to similar results across columns: an increase in ethnic identification by 0.011 (cultural groups), 0.013 (LEDA) or 0.009 (EPR), to be compared with 0.013 in our baseline.

Strength of an identity *versus* changing identity. Finally, note that all our estimations rest on the ethnic groups that the respondents declare. This declaration, i.e. the ethnic identity of the respondent, could itself react to natural resources. In particular, if natural resources benefit to certain groups, it might create some incentives for individuals to change their ethnic identities. Although ethnic identities are usually considered to be inherited at birth and quite stable, the literature has documented that "passing" may respond to economic incentives (Cornwell et al., 2017; Dahis et al., 2019), and that even ethnic identities can be prone to noisy signaling and manipulation in a lab setting (Harris et al., 2018). We check whether, additionally to the salience of the group identity, mineral resources affect the endogenous choice of ethnic identity, in the sense of passing from one group to another. We perform two different tests (Table A.20). First, we estimate how mining activity in the ethnic homeland of a group affects the population share of that group at the country-level. For each ethnic group×survey year, we compute the share of respondents that self-report belonging to that group. We regress this variable on the number of mines in the historical homeland of the group, as in our baseline estimations, also including ethnic group×country and country×year fixed effects. The estimates are close to zero and statistically insignificant across all four different definitions of ethnic groups (columns 1 to 4). Second, we estimate whether resources exploitation in an ethnic homeland affects the probability for the respondent to choose that ethnic group among all groups present in the country. We include the same set of fixed effects as in our baseline estimates. Column (5) shows a small negative estimate which is significant at the 10% level, but the number of observations is very large (about 1.8 million). Hence, overall, the results suggest that mining activity affects the strength of ethnic identities, but do not significantly causes changes from one ethnic identity to another. These results must however be interpreted with caution, as our data are not the most well-suited to estimate the determinants of changing identity.

4 How mining affects ethnic identity

We find that an increase in mineral resources exploitation makes feelings of ethnic identity stronger for the group whose historical homeland hosts the resource. In this section, we provide a set of

empirical exercises which globally support an interpretation of this finding rooted in the impact of natural resources on ethnic groups' grievances. At the core of our argument are the ethnic-group specific feelings of "relative deprivation". Inequality, real or perceived, generates discontent which increases the salience of identities (Olzak, 2011). This inequality may for example arise as the economic or political benefits of natural resource extraction are not disproportionately perceived to fall on local indigenous groups, or whenever natural resources are seen to cause a degradation of these groups' status or welfare. Building upon existing work and case studies presented below, our empirical approach considers the cycle between identity fragmentation, deprivation, and exclusion which may arise either in a decentralized manner or as a result from manipulation by "ethnic political entrepreneurs" (Aspinall, 2007).

4.1 Conceptual background and examples

Relative deprivation. At the individual level, feelings of relative deprivation – that is, the discrepancy between aspirations and reality – can be grounded either in inaccurate aspirations or in a detrimental economic reality. The start of industrial mining activities may raise aspirations related to economic development and resource rents (Collier, 2017), but under incomplete information, these aspirations may be unrelated to the reality of resources' extraction. Cust and Mihalyi (2017), for instance, show how natural resources discoveries trigger expectations unrelated to their economic reality, including among actors specialized in macroeconomic projections. At the individual level, Christensen (2019) argues that the link between mining activity and riots is rooted in incomplete information and (mis)perception about mining projects' profitability. This can be translated to the question of ethnic identities. Binzel and Carvalho (2017) show how identities can evolve in response to individual feelings of relative deprivation. Taking the example of the Egyptian revolution, they argue that economic development raises expectations, but does not always fulfill these. A religious identity then allows individuals to alter their reference point and to cope with these unfulfilled aspirations. They conclude that economic development may thus paradoxically make societies more prone to religious revivals. In our context, as mines' openings are taking place within the historical homeland of ethnic groups, we expect the discrepancy between aspirations and reality to be strongest for the members of these groups.[36] The aspirations and deprivation feelings of all the members of the resource rich ethnic groups may increase due to economic linkages, that would be amplified in case of a local multiplier effect (Moretti, 2010), or through an ethnic multiplier effect (being embedded in credit trade or migration networks with their co-ethnics, or through sheer in-group bias and solidarity - Fafchamps, 2000, 2003; Aker et al., 2014; Iwanowsky, 2018; Habyarimana et al., 2009). The Afrobarometer surveys include questions related to deprivation experiences and the perception of living conditions, which we use to test directly the grievances channel.

The role of leaders. Ethnic leaders, who range from traditional ethnic leaders – still a key local institution in Sub-Saharan Africa (Michalopoulos and Papaionnaou, 2020) – to national political representatives (Aspinall, 2007), may create or magnify individual feelings of relative deprivation

[36]Expectations should increase most among people who feel closest to the resource, here the members of the resource rich ethnic group. In a context where work identity is key, in Kazakhstan, Girard et al. (2021) show that expectations increase most among workers from the oil sector (compared to workers from other sectors) during an oil boom.

along ethnic lines. Such leaders may have an interest in doing so, to divert attention from their own rent extraction, or to use natural resource exploitation and ethnic discrimination in an opportunistic way in their political agenda, e.g. to gain popular support (Cheeseman and Larmer, 2015). The movement of Michael Sata in Zambia illustrates how a political leader exploits economic grievances related to mineral exploitation and articulates them along the lines of ethnic identity. Michael Sata practiced "ethnopopulism" (Cheeseman and Larmer, 2015), building upon the grievances of the Bemba identity group in the Zambian Copper belt.[37] We lack sufficient data to directly test this interpretation ; however, we note that, although it is quite indirect, our finding that the relation between resources and identity is magnified during electoral periods is consistent with "ethnic political entrepreneurs" creating or instrumenting ethnic-based grievances in their campaigns.

Natural resources, conflict, and ethnic identities. Further anecdotal evidence suggests that leaders may be particularly likely to successfully to capitalize over individual ethnic identity feelings in areas where feelings of ethnic inequality are pre-existing, such as in poor regions hosting resources, politically excluded ethnic groups, or in areas with a history of violent conflicts (Aspinall, 2007; Must, 2018). According to Aspinall (2007), who studies the Free Aceh movement in Indonesia, "rather than any intrinsic qualities of natural resource extraction, the key factor was the presence of an appropriate identity-based collective action frame." Aspinall argues that claims about the unjust exploitation of natural resources made by ethnic political entrepreneurs resonated powerfully in the population because they reinforced the long standing discourse of deprivation which infused Acehnese identity by the 1970s.[38] The qualitative work of Must (2018) suggests that similar forces were at play in Tanzania after large gas discoveries.[39] This literature suggests that leaders capitalization over ethnic identity may go as far as triggering ethnic conflicts, suggesting that identity exacerbation might be a missing link in the natural resources - conflict literature. The timing of the conflict-ethnic identity relationship is beyond the scope of this paper – one may follow the other, and both are likely to reinforce each other. However, this literature also suggests that pre-existing grievances and identity feelings, born from political exclusion, a history of conflict or long-standing poverty, may exacerbate the effect of natural resources on ethnic identity, in particular because these can be manipulated by ethnic political entrepreneurs. We take this prediction to the data at the end of this section, making use of information contained in the EPR dataset on political exclusion, and of data on local conflict and poverty. Section A14 of the appendix contains a more detailed description of the cases of Zambia, Indonesia and Tanzania.

[37] In the 2000s, "One consequence of the minerals boom was rising copper prices, which enabled Zambia's newly privatized mining companies to make vast profits, increasing public discontent with falling wages and living standards. [Michael Sata's] rhetoric chimed with an enduring perception that wealth produced by workers of Bemba origin was being illegitimately consumed by an incumbent ruling elite, but also with an ethnically diverse urban constituency in [the capital]." (Cheeseman and Larmer, 2015). See section A14 for more details on the country's background.

[38] Indonesia comprises different resource-rich provinces, (Aceh, Riau, East Kalimantan) which all experienced virtually identical processes of natural resource exploitation. However, the context of historical violence and institutionalization of ethnicity made Aceh distinct. As a result, only Aceh saw the development of ethnic tensions and intense separatism.

[39] Despite large regional marginalization, Tanzania remained peaceful for decades and an example of national integration (Green, 2013). Must (2018) insists on the role of leaders, perceived deprivation, and identities, concluding that "natural resource mismanagement and subsequent leadership framing increased the salience of a regional identity and exacerbated felt group grievances in southern Tanzania".

Competing channels and testable predictions. The main competing explanation of our findings is that mineral resources make ethnic identities more salient because they increase the payoffs associated with being part of a particular group. This could be the case if minerals have strong positive local economic effects (Aragón and Rud, 2013; Cust and Poelhekke, 2015; Benshaul-Tolonen, 2019; Mamo et al., 2019), that are felt disproportionately by historical local groups; or if they increase the political power and representation of these groups (Mamo and Bhattacharyya, 2018) or political patronage along ethnic lines (De Luca et al., 2018; Dickens, 2018). The actual payoffs of belonging to a given group would increase, which would make identities more salient. Combined with intra-ethnic spillovers due to migration, trade, remittances, or solidarity, this mechanism also explains why individuals residing outside their homeland also display stronger identity feelings. This "payoff" interpretation generates predictions which differ from the "deprivation" channel, and can be tested by looking at how minerals' extraction affect individual wealth, and whether the wealth of indigenous groups respond differently to mining.

4.2 Mining, wealth, and feelings of economic deprivation

Baseline. We first consider a number of Afrobarometer measures to study how individual wealth and *perceptions* of economic conditions react to mineral exploitation. To distinguish local effects from ethnic spillovers, the coefficients are estimated either on our baseline full sample, or on the sample of individuals residing outside their ethnic group's historical homeland. We proxy the objective material wealth of the household computing a "family of outcomes" index as the mean of the standardized variables using questions on whether the respondent personally owns a radio, a TV, and a vehicle (Kling et al., 2007).[40] Under the assumption that mineral resources' exploitation affects more (positively) the members of the ethnic group where mines are located, we would expect respondents to report owning more those goods. To measure economic deprivation, we build two different indicators. For the first, we compute a "family of outcomes" index based on five different questions in which individuals report whether their household has lacked over the past year essential welfare-related items: food, cash income, clean water, medicine or fuel to cook.[41] The second indicator uses information on individuals perceptions of changes in their living conditions. Individuals are asked about how their living conditions today compare to their living conditions 12 months before.[42] The last two variables have both an objective component – the individuals' actual wealth – and a subjective one, which reflect the individuals' perception of well-being compared to some reference point (see e.g. Ravallion, 2014). If resources exploitation do not alter the respondents' perceptions, we would expect all three variables to have a consistent

[40]The precise question is the following: "*Which of these things do you personally own: [Radio] / [TV] / [Vehicle]?*". Variables are standardized by subtracting their mean and dividing by their standard deviation. Though the resulting variable is only a proxy of actual wealth, it correlates well, as expected with weather shocks, as shown in section A8 of the appendix. Though the resulting variable is only a proxy of actual wealth, it does significantly correlate, as expected, with weather shocks (section A8 of the appendix).

[41]The precise question is the following: "*Over the past year, how often, if ever, have you or anyone in your family gone without:[Food] / [Cash income] / [Enough clean water for home use] / [Medicines or medical treatment] / [Enough fuel to cook your food?]* ". Answers are the following: 0=Never, 1=Just once or twice, 2=Several times, 3=Many times, 4=Always.

[42]The original question is the following: *"Looking back, how do you rate the following compared to twelve months ago: Your living conditions?"*. Answers are the following: 1=Much worse, 2=Worse, 3=Same, 4=Better, 5=Much better.

relation to resources exploitation. Results will however diverge if resources exploitation affect respondents' perceptions and heighten deprivation feelings independently of reality - as only the first variable is a purely objective measure of wealth.

Table 4: Natural resources extraction, wealth, and deprivation

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Wealth index		Deprivation index		Economic conditions	
Sample	Full	Outside homeland	Full	Outside homeland	Full	Outside homeland
# mines: ethnic homeland	0.006	0.004	0.004^{b}	0.008^{a}	-0.003	-0.008^{c}
	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)
Residence in homeland	-0.028^{a}		0.020^{b}		0.003	
	(0.006)		(0.009)		(0.011)	
Fixed effects	Country×Ethnic group, Region×Time					
Individual controls	Yes					
Observations	117775	63815	117788	63821	80667	43438
R^2	0.325	0.337	0.263	0.283	0.154	0.154
Sample mean / sd dep. var.	-0.02/0.71	0.11/0.72	-0.017/0.71	-0.026/0.72	2.95/1.05	2.99/1.035

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In columns (2), (4) and (6), we restrict the sample to individuals living outside their ethnic homeland. The dependent variable in columns (1) and (2) is an index of wealth based on ownership of radio, television and vehicles. In columns (3) and (4), it is the deprivation index, based on whether the respondent declares having gone without enough of the following items over the last year: food, income, water, medicine, fuel. Finally, in columns (5) and (6), the dependent variable is a categorical variable measuring the respondent's perception of his/her own living conditions compared to twelve months before. Even numbered columns restrict the sample to individual living outside their ethnic homeland. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

The results are provided in Table 4. Note that, contrary to the existing literature, these regressions do not estimate the overall effect of mining on wealth or welfare, given the presence of region × time fixed effects. Rather, they tell us whether the effects of mining are disproportionately felt by individuals belonging to the ethnic groups whose homeland hosts the mines – we come back to this issue in the next subsection. Several interesting findings arise. First, the effect of mineral resources on individual wealth is limited. The coefficient is positive, with a p-value of 0.102 in column (1), and both the magnitude and significance of the coefficient decrease in the sample of individuals residing outside their homeland (p-value=0.28, column 2). Quantitatively, an additional mine increases wealth by around 0.6%-0.8% of the standard deviation of the wealth index, which, to put things in perspective, represents between 3 and 4% of the effect of gender, and between 2 and 3% of the effect of being inactive (see Table A.21 in the appendix for the estimates of the control variables' coefficients). Second, results are more stable when using self-reports of economic deprivations and economic conditions (columns 3 to 6): in all columns individuals report worse outcomes, though the statistical significance varies in the case of perceptions of economic conditions. An additional mine implies a change equivalent to 0.6-1.1% of the standard deviation of the deprivation variable, i.e. between 30 and 72% of the effect of gender, and between 4 and 10% of the effect of being inactive. Third, though individuals living outside their homeland do not report significant changes in wealth (column 2), they report more deprivation (columns 4 and 6). Taken together, these results suggest that, for co-ethnics who do not live in their historical homeland, mines opening translates into larger deprivation feelings, possibly because those individuals, despite expecting benefits, do not enjoy the same direct benefit from the natural resources in comparison to individuals who live in their historical homeland.

Mines' characteristics. At this stage, these interpretations are slightly speculative, because our specification may be too rough to correctly identify the local economic effects of mining: the effects of mines may depend on technological characteristics (Pelzl and Poelhekke, 2021); it may also be very localized (Cust and Poelhekke, 2015), and vary non-linearly with the distance to the mines (De Haas and Poelhekke, 2019). Failing to consider these elements might lead our estimates to be noisy. To go further, we start by studying the role of mines' characteristics. Next, we consider a different specification which allows to measure more precisely the local effects of mines depending on the distance between the respondents and the resources exploitation.

Our data contains information on two types of characteristics that might affect the impact of mines on ethnic identities and economic spillovers. The first is the ownership structure of mines. Mines operated by large multinational companies might behave differently from large domestic or state-owned mines (e.g. Berman et al., 2017). Even if they do not, the very fact that a mine is domestic might affect how individuals perceive this mine and expect benefits form it. Second, our data also allows to proxy production technology and its labor/capital intensity. The recent literature has documented a Dutch disease type of effect of mineral production which depends on the technology of mining production (Pelzl and Poelhekke, 2021). Capital intensive mining tends to benefit more the non mining sectors because they cause no upward pressure on wages, contrary to labor-intensive mining which tends to be associated with a reduction in employment. Appendix A16 studies how these characteristics affect the link between mines and our various outcomes. We find a slightly stronger effect of domestic mines on the salience of ethnic identity. In addition, and consistent with Pelzl and Poelhekke (2021), only relatively capital intensive mines are associated with improvements in the wealth index. In both cases however, the coefficients are noisy, and overall we cannot conclude to statistically significant differences across mines types in most specifications. We interpret these findings as suggestive that the existence of local economic effects – and hence the payoffs associated with belonging to the local ethnic group – do not appear to be prime determinants of ethnic identification.

Mines' local effects. We go further by estimating the local effects of mines on wealth and deprivation feelings through a different econometric strategy. In the spirit of De Haas and Poelhekke (2019), our identification strategy exploits variations in the distance between the respondents and the mines. We first count the number of active mines located within a certain radius around each individual's residence, namely, [0 - 50 km] and [51 - 150 km].[43] We consider separately mines which, within this radius, are located in the ethnic homeland of the individual, and those which are not. We then estimate the impact of this set of mining activity variables on our proxies for wealth and economic deprivation, controlling for fixed effects at the country×ethnic group level, and at the country × time level. We use these aggregated fixed effects because our aim here is to estimate the average local effects of mines across all individuals (rather than to identify whether this effect varies across groups). As this approach prevents us to include the local time-varying

[43]The cutoff at 150 kilometers is similar to the one used by De Haas and Poelhekke (2019). We use a minimum radius of 50 kilometers around the mines instead of the 20 kilometers used by De Haas and Poelhekke (2019). 4% of the sample in Table 4 column 1 lives within 50 kilometers of a mine which is in their homeland, and 1% lives within 20 kilometers of such a mine. If we split by residence in the homeland to estimate the coefficients in column 2, 1.5% of respondents lives within 50 kilometers of a mine that is not in their homeland, and 0.2% of the sample lives within 20 kilometers of such a mine (87 individuals). We thus opt for cutoffs at 50 and 150 kilometers.

Table 5: Natural resources extraction: local effects

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Wealth index		Deprivation index		Economic conditions	
Sample	Full	Residents	Full	Residents	Full	Residents
# mines $[0-50]$ outside hom.	0.002	0.034^{b}	-0.002	-0.029^{c}	0.000	0.016
	(0.002)	(0.016)	(0.002)	(0.015)	(0.004)	(0.039)
# mines $[51-150]$ outside hom.	0.000	0.004	0.001	-0.005^{c}	0.002^{c}	0.011^{b}
	(0.000)	(0.005)	(0.001)	(0.003)	(0.001)	(0.005)
# mines $[0-50]$ in hom.	0.000	-0.008	0.006	0.014	-0.001	-0.003
	(0.006)	(0.006)	(0.009)	(0.012)	(0.007)	(0.011)
# mines $[51-150]$ in hom.	0.002	0.006	0.004^{a}	0.007	-0.002	-0.010
	(0.002)	(0.006)	(0.001)	(0.006)	(0.002)	(0.011)
Residence in homeland	-0.027^{a}		0.026^{b}		0.001	
	(0.006)		(0.012)		(0.015)	
Diff. in hom./outside hom.	-0.000	-0.040^{c}	0.011	0.055^{a}	-0.004	-0.040
	(0.008)	(0.022)	(0.009)	(0.019)	(0.008)	(0.029)
Fixed effects	Country×Ethnic group, Country×Time					
Individual controls	Yes					
Observations	117930	53915	117943	53922	80725	37169
R^2	0.295	0.295	0.198	0.202	0.103	0.124
Sample mean / sd dep. var.	-0.024/0.71	-0.067/0.7	-0.002 / 0.71	0.028/0.71	2.95 / 1.04	2.9/1.06

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In columns (2), (4) and (6), we restrict the sample to individual living inside their ethnic homeland. The dependent variable in columns (1) and (2) is an index of wealth based on ownership of radio, television and vehicles. In columns (3) and (4), it is the perceived deprivation index, based on whether the respondent declares having gone without enough of the following items over the last year: food, income, water, medicine, fuel. In columns (5) and (6), the dependent variable is a categorical variable measuring the respondent's perception of his/her own living conditions compared to twelve months before. Controls include age and its square, gender, rural/urban dummy, education categories, and employment status. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

fixed effects and makes endogeneity concerns stronger than in our baseline estimates, we also report results using mineral price variations to identify changes in mineral rents in appendix Table A.23.

Table 5 displays the estimated coefficients for the full sample of individuals (columns 1, 3 and 5) and for the sub-sample of individuals residing in the historical homeland of their ethnic group (columns 2, 4 and 6). The results confirm our previous interpretation: local wealth effects are limited (columns 1 and 2). Though most coefficients are positive, statistical significance is weak. More importantly, we fail to find any additional effect of resource extraction when the mines is within the homeland of the individual; jointly, the wealth effects appear significantly smaller in this case (column 2). Strikingly, while overall deprivation feelings seem to slightly decrease, and perceptions of economic conditions appear to improve in proximity to mines (columns 3 to 6), it is the opposite for mines located in the historical homeland of the individual. Put differently, mines increase deprivation feelings only for individuals belonging to resource-rich ethnic groups, those who host the mines in their historical homeland. These results are largely confirmed – and statistically reinforced – when using mineral price variations to identify changes in mineral rents (appendix Table A.23). Overall, these findings are consistent with the idea that resources windfalls

do not appear to be felt by the ethnic group whose historical homeland hosts the resource. Rather, resource rich individuals may feel economically deprived, in particular individuals living in their historical homeland – or leaders in these homelands may use mines to foster such feelings – and these perceptions might be shared by or transmitted to co-ethnics living outside these homelands.

4.3 Political exclusion and ethnic-group characteristics

As discussed in Section 4.1, though the mechanisms discussed above can in principle apply to many contexts, we expect them to be particularly relevant when feelings of ethnic inequality are pre-existing. This would be the case for politically excluded ethnic groups, for groups with a strong pre-existing sense of ethnic identity, and for those originating from (relatively) poor areas, or areas with a history of violent conflict (Aspinall, 2007; Must, 2018; Vogt, 2017). Differences across ethnic groups along these dimensions may contribute to explain why we find a substantial amount of heterogeneity in the link between mining and ethnic identification (Figure A.8). In this section, we take these predictions to the data.

Political exclusion. We start by investigating the role of political exclusion (Table 6). We make use of specific information about the political power of ethnic groups from the Ethnic Power Relations Dataset (EPR, Wucherpfennig et al., 2011; Vogt et al., 2015), which records contemporaneous ethnic homeland boundaries as well as time-varying political power. The aim of the EPR data is to focus on ethnic groups who have a clear political relevance in their countries, leaving many groups outside its sample.[44] As a preliminary check, we thus replicate the results of the baseline Table 2, after restricting the sample to groups which *do not* appear in the EPR sample. The magnitude of the estimates of mineral resources extraction in an ethnic homeland is systematically higher (more than three times) in the restricted sample than in the baseline sample (Table A.24 in the Appendix). In other words, the effect of mineral resources on ethnic identification appears to be drastically magnified in the sample of groups without any ethnic political representation.

We then exploit the information present in the EPR dataset. We interact the mining variable with dummies capturing the degree of political power of the group, allowing heterogeneous effects by group political power (compared to the aggregate effect presented in the appendix Table A.18). While all groups of the EPR sample are politically relevant, they differ in their ability to influence political decisions. We split the groups into two mutually exclusive categories: powerful groups, which have some access to power, and powerless ones, which do not.[45] Because the EPR status

[44] It may even be entire countries which do not appear in the EPR data. Vogt et al. (2015), p. 4, note that "An ethnic group is considered politically relevant if at least one political organization has claimed to represent its interests at the national level or if its members are subjected to state-led political discrimination (Cederman, Wimmer, and Min 2010, 99)." Group identities may be salient in everyday lives, like in wedding or business networks, while still not having a formal political representation. In Burkina Faso for example, the share of inter-ethnic marriages among married couples in 2003 was only 11.5%, while this rate would have been 67.5% if marriage matches were orthogonal to ethnic group identities (Crespin-Boucaud, 2020, appendix Table 2). Yet, Burkina Faso is a country where the EPR dataset does not record a single ethnic group.

[45] We include the following EPR categories in the groups with access to power: dominant, junior partner, senior partner. 87% of the Afrobarometer respondents belong to a group with access to power (30 groups) while 13% of the respondents belong to a powerless group (5 groups). The powerless groups contain the following EPR categories: discriminated, irrelevant, powerless. The other categories (e.g. monopoly, self-exclusion) are never present in our sample. The EPR categories definition are as follows. "1. The group rules alone: monopoly or dominant. In contrast to monopoly power, the status of dominant indicates 'token' representation of other ethnic groups in the executive.

Table 6: Natural resources extraction and political exclusion

Dep. Var.	(1) Ethnic identification	(2) Group Treated unfairly	(3) Wealth index	(4) Deprivation index	(5) Economic conditions
# mines: ethnic homeland × powerful	0.009	0.022[a]	-0.003	0.000	-0.006
	(0.008)	(0.006)	(0.002)	(0.002)	(0.006)
# mines: ethnic homeland × powerless	0.186[a]	0.331[a]	-0.014	0.033	-0.059[b]
	(0.021)	(0.065)	(0.015)	(0.022)	(0.028)
Resids in homeland	0.056[a]	0.064[a]	-0.011	0.020[b]	-0.002
	(0.018)	(0.016)	(0.007)	(0.009)	(0.013)
Fixed effects	Country×Ethnic group, Region×Time				
Individual controls	Yes				
Observations	84841	81138	85859	85865	58492
R^2	0.177	0.228	0.347	0.284	0.147
Sample mean / sd dep. var.	2.35/1.2	2.36/1.2	0.04/0.74	-0.06/0.72	2.99/1.06
Test differences (p-value)	0.000	0.000	0.433	0.132	0.046

OLS estimations. [c] significant at 10%; [b] at 5%; [a] at 1%. Standard errors are clustered at the ethnic group level. The dependent variable in column (1) is the level of ethnic identification with respect to state identification in all columns, ranging from categories 1 to 5. In column (2), the dependent variable tells if the respondents considers that her ethnic group has been treated unfairly by the government, ranging from categories 0 (never) to 3 (always). The dependent variable in column (3) is an index of wealth based on ownership of radio, television and vehicles. In column (4), it is the perceived deprivation index, based on whether the respondent declares having gone without enough of the following items over the last year: food, income, water, medicine, fuel. Finally, in column (5), the dependent variable is a categorical variable measuring the respondent's perception of his/her own living conditions compared to twelve months before. Ethnic homeland and ethnic groups are based on data from EPR. Powerful groups which fall into one of the following EPR categories in the EPR dataset at the beginning of the period: dominant, junior and senior partners. Powerless are groups will fall in the following categories: discriminated, irrelevant, powerless. We separately control for mines abroad. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. Test differences (p-value) tells the p-value of the test that the effect of the number of mines in the homeland of a powerful group is equal to the effect of the number of mines in the homeland of a powerless group. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

may be endogenous to variations in mineral resource production, we use the political status of groups before the start of our study period, in 2004. We find that hosting mineral resources production has no significant effect on the level of ethnic identification of groups with access to power (column 1). On the other hand, the effect is large and significant for powerless groups – an order of magnitude higher than in our baseline estimates.[46]

In column (2), we assess more directly the impact of mining activity on the sharing of power across ethnic groups, as perceived by local households. We make use of an Afrobarometer question in which individuals are asked whether they believe that their ethnic group is treated unfairly by the government.[47] We find a positive and significant effect of mining activity on perceptions of unequal ethnic group treatment, which is again quantitatively much stronger for individuals belonging to powerless groups. We interpret this result as a strong signal that mineral exploitation increases identity-based feelings of deprivation, be these feelings linked to pure economic motives,

2. The group shares power: senior partner or junior partner, depending on the group's absolute influence in the executive (i.e. irrespective of group size). 3. The group is excluded: powerless, discriminated, or self-exclusion. While powerless means that the group is simply not represented (or does not have influence) in the executive, discrimination indicates an active, intentional, and targeted discrimination by the state against group members in the domain of public politics." (Vogt et al., 2015, p. 6-7)

[46]Note that we split the effect for the number of mines in the ethnic homeland in the country of the respondent, because in our sample mines located abroad are systematically in the homelands of groups who have access to political power. This lack of variation prevents us from identifying the heterogeneous effect of mines abroad by access to power. However, it is interesting to note that the coefficients have closer magnitude for resources abroad and resources in the homeland of a powerful group, than for resources in the homeland of a powerless group.

[47]The question is stated as: *"How often is [Respondent?s Ethnic Group] treated unfairly by the government"*. Answers are the following: 0=Never, 1=Sometimes, 2=Often, 3=Always

or to a drop in the trust that individuals have in their national government.

Finally, in columns (3) to (5), we show that the effects found in the previous section on economic wealth and deprivation indexes are reinforced in the case of powerless groups. Though neither powerless nor powerful groups' wealth appear to improve with mining activity (column 3), powerless groups exhibit stronger increases in deprivation (column 4; p-value = 0.13) and more pessimism about economic conditions (column 5). The extraction of mineral resources in the homeland of politically powerful ethnic groups either has a smaller and always statistically insignificant effect on the various outcomes.

Conflict and poverty. Finally, we consider the exacerbating role of poverty, conflict and pre-existing ethnic feelings (Table 7). We construct three ethnic-group specific measures. The first investigates the consequence of pre-existing conflicts.[48] We use information from the Armed Conflict Location Events Data (ACLED, Raleigh et al., 2010) on the location, date and types of conflict events to build a control for the level of conflicts taking place in each homeland and year. For each ethnic homeland, we compute the number of battle events and other conflict events recorded in ACLED over the period 2000-2004 (before our sample starts). Second, we proxy development level using nighttime luminosity data, which have been widely used by recent literature as a proxy for local income (Donaldson and Storeygard, 2016). The data come from the National Oceanic and Atmospheric Administration. For each homeland, we compute the average total luminosity over the period 2000-2004, relative to the country's. Finally, we measure overall pre-existing ethnic feelings by computing, for each homeland, the average ethnic identity in the first wave of the Afrobarometer during which the ethnic group is surveyed and define homelands with strong pre-existing feelings of ethnic identity as homelands where this average is above the median.

The results point out that groups most likely to have pre-existing feelings of ethnic inequality are those who react the most to mineral resources exploitation in their historical ethnic homeland (Table 7). We first show that violent conflict (battles) history in the homeland magnifies our coefficient of interest, while other type of conflict events have little or no effect, be it in the full sample or focusing on households living outside of their historical homeland (columns 1, 3 and 4). On the other hand, a higher pre-period income attenuates our baseline coefficient, i.e. respondents whose ethnic homelands are poorer are the one who react the most, in terms of ethnic identification, when natural resources' exploitation intensifies in their homeland (columns 2 and 3), although this effect loses statistical significance for individuals residing outside their homeland (column 4). Finally, pre-existing ethnic group specific levels of ethnic identification matter. Groups with stronger pre-existing identification feelings (measured in the first survey wave) identify more to their ethnic groups when mineral exploitation takes place in their homeland, while respondents from groups with low pre-existing levels of identification do not react to the presence of natural resources. These various results are consistent with the channel of relative deprivation and unequal treatment.

[48]The timing of the relation between conflict and identity is subject to debate as conflicts may exacerbate existing identities (Atkin et al., 2019), but existing identities can also be used to mobilize conflict participants (Esteban et al., 2012). We acknowledge that both contemporaneous conflicts and contemporaneous identities may be jointly determined, potentially one and only thing and prevent us to distinguish the timing or the relevance of each force within our baseline results. Rather, what we do here is to focus on how pre-existing conflicts and other pre-existing conditions related to relative deprivation interact with mineral exploitations.

Table 7: Ethnic-group specific factors

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Salience of ethnic identification					
Sample	——Full——			Outside homeland	Full	Outside homeland
# mines: ethnic homeland	0.018^{a}	0.029^{a}	0.027^{a}	0.017^{a}	-0.002	-0.006
	(0.003)	(0.005)	(0.005)	(0.006)	(0.012)	(0.013)
× # battles$_{t0}$	0.016^{a}		0.012^{b}	0.018^{a}		
	(0.006)		(0.006)	(0.006)		
× # other conflicts$_{t0}$	-0.000^{a}		-0.000	-0.000		
	(0.000)		(0.000)	(0.000)		
× Nighttime lights$_{t0}$		-0.009^{a}	-0.008^{c}	-0.001		
		(0.003)	(0.004)	(0.005)		
× strong ethnic identification$_{t0}$					0.035^{b}	0.029^{c}
					(0.016)	(0.017)
Residence in homeland	0.045^{a}	0.045^{a}	0.045^{a}		0.040^{a}	
	(0.012)	(0.012)	(0.012)		(0.014)	
Fixed effects	Country×Ethnic group, Region×Time					
Individual controls	Yes					
Observations	115105	115105	115105	61790	76103	39697
R^2	0.185	0.185	0.185	0.180	0.181	0.178
Sample mean / sd dep. var.		2.35/1.19		2.31/1.18	2.28/1.18	2.23/1.17

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In columns (4) and (6), we restrict the sample to individual living outside their ethnic homeland. The dependent variable is the level of ethnic identification with respect to state identification in all columns, ranging from categories 1 to 5. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. We interact the number of active mines in the homeland with variables accounting for the pre-existing level of battles, other conflicts, nightlights and identification. # battles$_{t0}$ (respectively # other conflict$_{t0}$) is the number of battles (resp. other conflict events), in the ethnic homeland of the individual, over the period 2000-2004 (ACLED). Nighttime lights$_{t0}$ is the average level of nighttime lights over 2000-2004. strong ethnic identification$_{t0}$ is a dummy variable taking value one for groups where the average level of ethnic identification in the first Afrobarometer wave is above the sample median (this first wave is excluded from the sample in columns (5) and (6), we interact this group-level dummy with the number of active mines). Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

5 Conclusion

While we know the risks of fragmented identities, we still know little about how one identity becomes so central that it shapes our actions and lives. Here, we document why natural resources may be an important driver of identities fragmentation in Sub-Saharan Africa. We exploit geo-localized individual data on the strength of ethnic *versus* national identification in 25 countries of sub-Saharan Africa from 2005 to 2015, matched to the respondents' ethnic homelands, and mineral resources exploitation in these homelands.

Our results suggest that mineral resources fosters identity fragmentation. We document a significant increase in the strength of ethnic identification, relative to national identification, as a response to the exploitation of mineral resources in the ethnic group's historical homeland. The effect is persistent, and tends to be magnified during elections periods, which implies that changes in resource exploitation might have a long-lasting impact on ethnic identification through voting patterns. We argue that the mechanism underlying this finding is one of economic deprivation, ethnic grievances and political exclusion. We fail at identifying substantial local economic

spillovers from mining activity in our sample and any significant additional economic benefit for individuals belonging to the group whose historical homeland hosts mining activities. On the other hand, the latter individuals report more economic deprivation and pessimism about economic conditions. We find that our results are mainly driven by individuals belonging to politically excluded groups. We show that pre-existing feelings of ethnic inequality, through political exclusion, conflict, or poverty, matter. An open question is whether such grievances arise in a decentralized manner or result from manipulation by "ethnic political entrepreneurs" – an interesting extension would be to study leaders' speeches to answer this question.

Overall, this paper documents a potentially understudied consequence of natural resources extraction, and at the same time sheds new light on the sources of identities fragmentation, an issue which has potential implications for inter-group conflict as well as economic performance. Our results suggest that perceptions of economic or political deprivation can play a substantial role in driving identity constructions, regardless of actual economic achievements.

References

Ahlerup, P., T. Baskaran, and A. Bigsten (2017). Regional development and national identity in sub-Saharan Africa. *Journal of Comparative Economics 45*(3), 622–643.

Aker, J. C., M. W. Klein, S. A. O'Connell, and M. Yang (2014). Borders, ethnicity and trade. *Journal of Development Economics 107*(C), 1–16.

Akerlof, G. A. and R. E. Kranton (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press.

Alesina, A., R. Baqir, and W. Easterly (1999, 11). Public Goods and Ethnic Divisions. *The Quarterly Journal of Economics 114*(4), 1243–1284.

Alesina, A. and E. La Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature 43*(3), 762–800.

Alsan, M. (2015). The Effect of the TseTse Fly on African Development. *American Economic Review 105*(1), 382–410.

Anderson, B. (2006). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso books.

Aragón, F. M., P. Chuhan-Pole, and B. C. Land (2015). *The local economic impacts of resource abundance: What have we learned?* The World Bank.

Aragón, F. M. and J. P. Rud (2013). Natural resources and local communities: Evidence from a Peruvian gold mine. *American Economic Journal: Economic Policy 5*(2), 1–25.

Arezki, R., V. A. Ramey, and L. Sheng (2017). News shocks in open economies: Evidence from giant oil discoveries. *The quarterly journal of economics 132*(1), 103–155.

Armand, A., A. Coutts, P. C. Vicente, and I. Vilela (2020). Does information break the political resource curse? experimental evidence from mozambique. *American Economic Review 110*(11), 3431–53.

Asiwaju, A. (1985). Partitioned africans. the conceptual framework.

Aspinall, E. (2007). The construction of grievance: Natural resources and identity in a separatist conflict. *Journal of Conflict Resolution 51*(6), 950–972.

Atkin, D., E. Colson-Sihra, and M. Shayo (2019, March). How do we choose our identity? a revealed preference approach using food consumption. *National Bureau of Economic Research Working Paper Series*.

Barrios, S., L. Bertinelli, and E. Strobl (2010). Trends in rainfall and economic growth in africa: A neglected cause of the african growth tragedy. *The Review of Economics and Statistics 92*(2), 350–366.

Benshaul-Tolonen, A. (2019). Local industrial shocks and infant mortality. *The Economic Journal 129*(620), 1561–1592.

BenYishay, A., R. Rotberg, J. Wells, Z. Lv, S. Goodman, L. Kovacevic, and D. Runfola (2017). Geocoding Afrobarometer Rounds 1-6: Methodology & Data Quality. *Williamsburg VA: AidData at William & Mary*.

Berman, N., M. Couttenier, D. Rohner, and M. Thoenig (2017). This Mine Is Mine! How Minerals Fuel Conflicts in Africa. *American Economic Review 107*(6), 1564–1610.

Binzel, C. and J.-P. Carvalho (2017). Education, social mobility and religious movements: The islamic revival in egypt. *The Economic Journal 127*(607), 2553–2580.

Bisin, A., E. Patacchini, T. Verdier, and Y. Zenou (2016). Bend it like Beckham: Ethnic identity and integration. *European Economic Review 90*(C), 146–164.

Caselli, F. and G. Michaels (2013). Do Oil Windfalls Improve Living Standards? Evidence from Brazil. *American Economic Journal: Applied Economics 5*(1), 208–38.

Cheeseman, N. and M. Larmer (2015). Ethnopopulism in africa: Opposition mobilization in diverse and unequal societies. *Democratization 22*(1), 22–50.

Christensen, D. (2019). Concession Stands: How Mining Investments Incite Protest in Africa. *International Organization 73*(1), 65–101.

Colella, F., R. Lalive, S. Sakalli, and M. Thoenig (2020). Acreg: Stata module to perform arbitrary correlation regression.

Collier, P. (2017). The institutional and psychological foundations of natural resource policies. *The Journal of Development Studies 53*(2), 217–228.

Conley, T. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*, 1–45.

Cornwell, C., J. Rivera, and I. Schmutte (2017, 06). Wage discrimination when identity is subjective: Evidence from changes in employer-reported race. *Journal of Human Resources 52*, 719–755.

Crespin-Boucaud, J. (2020). Interethnic and interfaith marriages in sub-Saharan Africa. *World Development 125*, 104668.

Cust, J. and J. Mensah (2020). Natural resource discoveries, citizen expectations and decisions.

Cust, J. and D. Mihalyi (2017). Evidence for a presource curse? Oil discoveries, elevated expectations, and growth disappointments. *World Bank Working Paper Series*.

Cust, J. and S. Poelhekke (2015, October). The Local Economic Impacts of Natural Resource Extraction. *Annual Review of Resource Economics 7*(1), 251–268.

Dahis, R., Emily, and N. Qian (2019). Choosing Racial Identity in the United States, 1880-1940. *National Bureau of Economic Research Working Paper Series* (26465).

De Haas, R. and S. Poelhekke (2019). Mining matters: Natural resource extraction and firm-level constraints. *Journal of International Economics 117*, 109–124.

De Luca, G., R. Hodler, P. A. Raschky, and M. Valsecchi (2018). Ethnic favoritism: An axiom of politics? *Journal of Development Economics 132*, 115–129.

Dehdari, S. H. and K. Gehring (2022, January). The origins of common identity: Evidence from alsace-lorraine. *American Economic Journal: Applied Economics 14*(1), 261–92.

Dell, M., B. F. Jones, and B. A. Olken (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics 4*(3), 66–95.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature 52*(3), 740–98.

Depetris-Chauvin, E., R. Durante, and F. R. Campante (2020). Building Nations Through Shared Experiences: Evidence from African Football. *The American Economic Review 10*(5), 1572–1602.

Dickens, A. (2018). Ethnolinguistic Favoritism in African Politics. *American Economic Journal: Applied Economics 10*(3), 370–402.

Donaldson, D. and A. Storeygard (2016). The View from Above: Applications of Satellite Data in Economics. *Journal of Economic Perspective*, 171–98.

Easterly, W. and R. Levine (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. *The Quarterly Journal of Economics, MIT Press 112*(4), 1203–50.

Eifert, B., E. Miguel, and D. N. Posner (2010). Political Competition and Ethnic Identification in Africa. *American Journal of Political Science 54*(2), 494–510.

Esteban, J., L. Mayoral, and D. Ray (2012). Ethnicity and conflict: An empirical study. *American Economic Review 102*(4), 1310–42.

Esteban, J. and D. Ray (2011). Linking conflict to inequality and polarization. *American Economic Review 101*(4), 1345–74.

Fafchamps, M. (2000). Ethnicity and credit in african manufacturing. *Journal of Development economics 61*(1), 205–235.

Fafchamps, M. (2003). Ethnicity and networks in african trade. *Contributions in Economic Analysis & Policy 2*(1), 1–51.

Fenske, J. and I. Zurimendi (2017). Oil and ethnic inequality in Nigeria. *Journal of Economic Growth 22*(4), 397–420.

Gehring, K. and S. A. Schneider (2020). Regional resources and democratic secessionism. *Journal of Public Economics 181*, 104073.

Gennaioli, N. and I. Rainer (2007). The modern impact of precolonial centralization in Africa. *Journal of Economic Growth 12*(3), 185–234.

Girard, V., A. Kudebayeva, and G. Toews (2021). Inflated expectations and commodity prices: Evidence from kazakhstan. In *Handbook of Sustainable Politics and Economics of Natural Resources*. Edward Elgar Publishing.

Green, E. (2013). The political economy of nation formation in modern tanzania: explaining stability in the face of diversity. In *Nationalism and Conflict Management*, pp. 77–98. Routledge.

Green, E. (2018). Ethnicity, National Identity and the State: Evidence from Sub-Saharan Africa. *British Journal of Political Science*, 1–23.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review 101*(4), 709–725.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2009). *Coethnicity: diversity and the dilemmas of collective action*. Russell Sage Foundation.

Harris, A., D. L. Nielson, L. Medina, W. Berlin, C. B. M. Correia, M. G. Findley, J. M. Weinstein, J. Habyarimana, M. Humphreys, and D. N. Posner (2018). Experimental Evidence from Uganda, South Africa, and the United States on Ethnic Identification and Ethnic Deception. *Prepared for the APSA Annual Conference, Boston*.

Ian Harris, Timothy J. Osborn, P. J. and D. Lister (2020). Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset. *Scientific data - Nature 190*(7).

Iwanowsky, M. (2018). The effects of migration and ethnicity on african economic development. *Unpublished Mansucript available at http://www. mathiasiwanowsky. com/Migration_and_Trade. pdf*.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Klor, E. and M. Shayo (2010). Social identity and preferences over redistribution. *Journal of Public Economics 94*(3-4), 269–278.

Larmer, M. (2016). *At the Crossroads: Mining and Political Change on the Katangese-Zambian Copperbelt*. Oxford University Press.

Lessmann, C. and A. Steinkraus (2019). The geography of natural resources, ethnic inequality and civil conflicts. *European Journal of Political Economy 59*, 33–51.

Loayza, N. and J. Rigolini (2016). The local impact of mining on poverty and inequality: evidence from the commodity boom in Peru. *World Development 84*, 219–234.

Lowes, S., N. Nunn, J. A. Robinson, and J. Weigel (2015). Understanding ethnic identity in africa: Evidence from the implicit association test (iat). *American Economic Review 105*(5), 340–45.

Lungu, J. (2008). Copper mining agreements in zambia: renegotiation or law reform? *Review of African Political Economy 35*(117), 403–415.

Mamo, N. and S. Bhattacharyya (2018). Natural Resources and Political Patronage in Africa: An Ethnicity Level Analysis. *Department of Economics, University of Sussex Business School* (0418).

Mamo, N., S. Bhattacharyya, and A. Moradi (2019). Intensive and extensive margins of mining and development: evidence from Sub-Saharan Africa. *Journal of Development Economics 139*, 28–49.

Michalopoulos, S. and E. Papaioannou (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica 81*(1), 113–152.

Michalopoulos, S. and E. Papaionnaou (2020). Historical legacies and African Development. *Journal of Economic Literature*.

Miguel, E. (2004). Tribe or nation? nation building and public goods in kenya versus tanzania. *World Politics 56*(3), 327–362.

Monteiro, J. and C. Ferraz (2010). Does oil make leaders unaccountable? evidence from brazil's offshore oil boom. *unpublished, PUC-Rio*.

Morelli, M. and D. Rohner (2015). Resource concentration and civil wars. *Journal of Development Economics 117*, 32 – 47.

Moretti, E. (2010). Local multipliers. *American Economic Review 100*(2), 373–77.

Muller-Crepon, C., Y. I. Pengl, and N.-C. Bormann (2020). Linking ethnic data from africa. *Journal of Peace Research*, forthcoming.

Murdock, G. P. (1959). An Atlas of African History. JD Fage. *American Anthropologist 61*(3), 530–531.

Must, E. (2018). Structural inequality, natural resources and mobilization in southern tanzania. *African Affairs 117*(466), 83–108.

Must, E. and S. A. Rustad (2016). Perceptions of justice and violent mobilization: Explaining petroleum related riots in southern tanzania. *Prio Conflict Trends 06*.

Must, E. and S. A. Rustad (2019). "mtwara will be the new dubai": Dashed expectations, grievances and civil unrest in tanzania. *International Interactions 45(3)*, 500–531.

Ndulu, B. J., W. E. Mbowe, and E. Hunter (2019). *Ethnicity, Citizenry, and Nation-Building in Tanzania*, pp. 98–122. Cambridge University Press.

Nunn, N. (2008). The Long-term Effects of Africa's Slave Trades. *The Quarterly Journal of Economics 123*(1), 139–176.

Nunn, N. and L. Wantchekon (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review 101*(7), 3221–52.

Olzak, S. (2011). Does globalization breed ethnic discontent? *Journal of Conflict Resolution 55*(1), 3–32.

Pelzl, P. and S. Poelhekke (2021). Good mine, bad mine: Natural resource heterogeneity and dutch disease in indonesia. *Journal of International Economics 131*.

Pengl, Y. I., P. Roessler, and V. Rueda (2021). Cash crops, print technologies, and the politicization of ethnicity in africa. *American Political Science Review*, 1–19.

Posner, D. N. (2004). The political salience of cultural difference: Why chewas and tumbukas are allies in zambia and adversaries in malawi. *American Political Science Review*, 529–545.

Raleigh, C., A. Linke, H. Hegre, and J. Karlsen (2010). Introducing acled: an armed conflict location and event dataset: special data feature. *Journal of Peace Research 47*(5), 651–660.

Ravallion, M. (2014). Poor, or just feeling poor? on using subjective data in measuring poverty. *Happiness and Economic Growth*, 140–178.

Robinson, A. L. (2014). National versus ethnic identification in Africa: Modernization, colonial legacy, and the origins of territorial nationalism. *World Politics 66*(4), 709–746.

Rohner, D., M. Thoenig, and F. Zilibotti (2013). Seeds of distrust: conflict in Uganda. *Journal of Economic Growth 18*(3), 217–252.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review 103*(2), 147–174.

van der Ploeg, F. (2011, June). Natural Resources: Curse or Blessing? *Journal of Economic Literature 49*(2), 366–420.

Venables, A. J. (2016, February). Using natural resources for development: Why has it proven so difficult? *Journal of Economic Perspectives 30*(1), 161–84.

Vogt, M. (2017). Escaping the resource curse: Ethnic inclusion in resource-rich states in west africa. In *Issues in African Political Economies*, pp. 199–225. Carolina Academic Press.

Vogt, M., N.-C. Bormann, S. Rüegger, L.-E. Cederman, P. Hunziker, and L. Girardin (2015). Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution 59*(7), 1327–1342.

Wucherpfennig, J., N. B. Weidmann, L. Girardin, L.-E. Cederman, and A. Wimmer (2011). Politically relevant ethnic groups across space and time: Introducing the geoepr dataset. *Conflict Management and Peace Science 28*(5), 423–437.

Appendix

A1 Afrobarometer sample

The years of survey for yeach of the 25 countries included in the sample are: Benin (2005, 2008, 2011, 2014), Botswana (2005, 2008, 2012, 2014), Burkina Faso (2008, 2012, 2015), Cameroon (2013, 2015), Cote d'Ivoire (2013, 2014), Ghana (2005, 2008, 2012, 2014), Guinea (2013, 2015), Kenya (2005, 2008, 2011, 2014), Lesotho (2005, 2008, 2012, 2014), Liberia (2008, 2012, 2015), Madagascar (2005, 2008, 2013, 2014, 2015), Malawi (2005, 2008, 2012, 2014), Mali (2005, 2008, 2012, 2013, 2014), Mozambique (2005, 2008, 2012, 2015), Namibia (2006, 2008, 2012, 2014), Niger (2013, 3015), Nigeria (2005, 2008, 2012, 2013, 2014, 2015), Senegal (2005, 2008, 2013, 2014), Sierra Leone (2012, 2015), South Africa (2011, 2015), Tanzania (2005, 2008, 2012, 2014), Togo (2012, 2014), Uganda (2005, 2008, 2011, 2012, 2015), Zambia (2009, 2011, 2012, 2015), Zimbabwe (2012, 2014).

A2 Matching Afrobarometer ethnic groups to Murdock homelands

We pursue different strategies to match the ethnic groups recorded in the Afrobarometer surveys and in the Murdock Atlas. The list of ethnic groups recorded in the Afrobarometer is directly inspired from the Murdock Atlas. However, the Afrobarometer also updates that list. As a result, the Afrobarometer acknowledges that some homelands recorded in Murdock have given rise to different groups today. For example, the Bafokeng, Mofokeng, Mkwere, Kikwere and Wakwere ethnic groups that the Afrobarometer records in Lesotho and Tanzania correspond to a single ethnic homeland from Murdock, that of the Kwena. Conversely, some of the ethnic homelands recorded distinctly in Murdock shared proximity such that it makes sense to record them as a single entity in the Afrobarometer. In Ivory coast for example, the Afrobarometer questionnaire records only five ethnic groups, while the Murdock Atlas records 35 ethnic groups.

We exploit three main sources to match the groups recorded in the Afrobarometer and the Murdock Atlas (we used the same for the matching to the EPR dataset): extracts and trees

of relations between groups from the Murdock Atlas,[49] the matching made by Nunn (2008) for countries of the survey round 3, and the Ethnologue.[50] In the cases when none of the above was enough, we crossed information from the Joshua Project,[51] Wikipedia entries on the ethnic group name,[52] and press articles. In a last step, we cross checked our matching against the matching proposed by the LEDA (Muller-Crepon et al., 2020) and described below. Whenever one of the matches proposed by the LEDA overlapped with our initial match, we kept that match. Whenever there was no overlap, we performed extra searches before settling which match to keep.

We check the robustness of our results to two alternative definitions of ethnic homelands in the Murdock Atlas. In the first, we aim at taking in account all the Murdock homeland which may relate to an identity group from the Afrobarometer. To do so, we split respondents between Cultural areas rather than Murdock homelands as the Murdock Atlas attributes each ethnic group to a broader Cultural group. Cultural groups thus cover bigger geographic areas than Murdock homelands: the 296 Murdock ethnic groups we record belong to only 71 distinct Cultural areas. However, as appears in Figure A.3 cultural groups allow to cover virtually all the relevant geographic areas (expanding in space our baseline definition) without leaving any endogeneity in the choice of which group stands alone and which group gets merged to others.

In the second, we make use of the recently released tool of the LEDA: Linking Ethnic Data from Africa (Muller-Crepon et al., 2020). The LEDA restricts the definition of ethnicity to linguistic identity categories, explicitly leaving aside the historical evolution or subjective beliefs in common descent. Doing so allows the authors to exploit the linguistic tree from the Ethnologue database. (Muller-Crepon et al., 2020) link to this language tree the ethnic groups recoded in different databases (including the Afrobarometer and Murdock databases), and then use the position of the groups in the language tree to propose a link between them. We proceed in several steps to exploit this data. First, we match groups based on their set relation (telling the share of branches where they overlap in the language tree). Second, we restrict the sample to the matches that ensure a maximum possible overlap of branches in the language tree. Third, we drop matches for which the maximum overlap is below 0.6.[53] Last, when several such groups exist, we randomly restrict the match to one of the groups proposed.Overall, this procedure match the Afrobarometer respondents to 413 Murdock homelands. Results appear in Figure A.4. Our baseline matching and the LEDA matching result in a similar match for 280 groups, which represent 63% of the sample of respondents.

[49] https://www.webafriqa.net/library/anthropology/murdock/

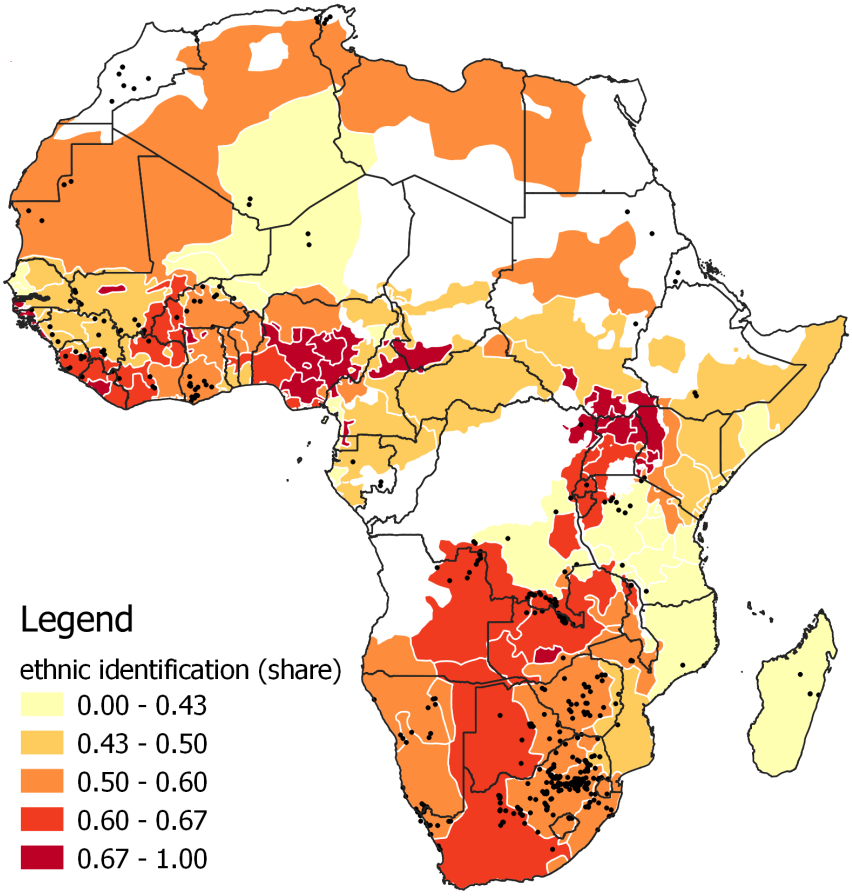
[50] https://www.ethnologue.com/

[51] https://joshuaproject.net/ which shows a contemporaneous catalog of ethnic groups and their languages compiled by a missionary group, as was the initial trigger of the Ethnologue catalog.

[52] Mainly to check the translation of the group names in different languages by changing the language of the Wikipedia entry. Wikipedia also provides information on synonymous and related groups for example in https://en.wikipedia.org/wiki/List_of_contemporary_ethnic_groups.

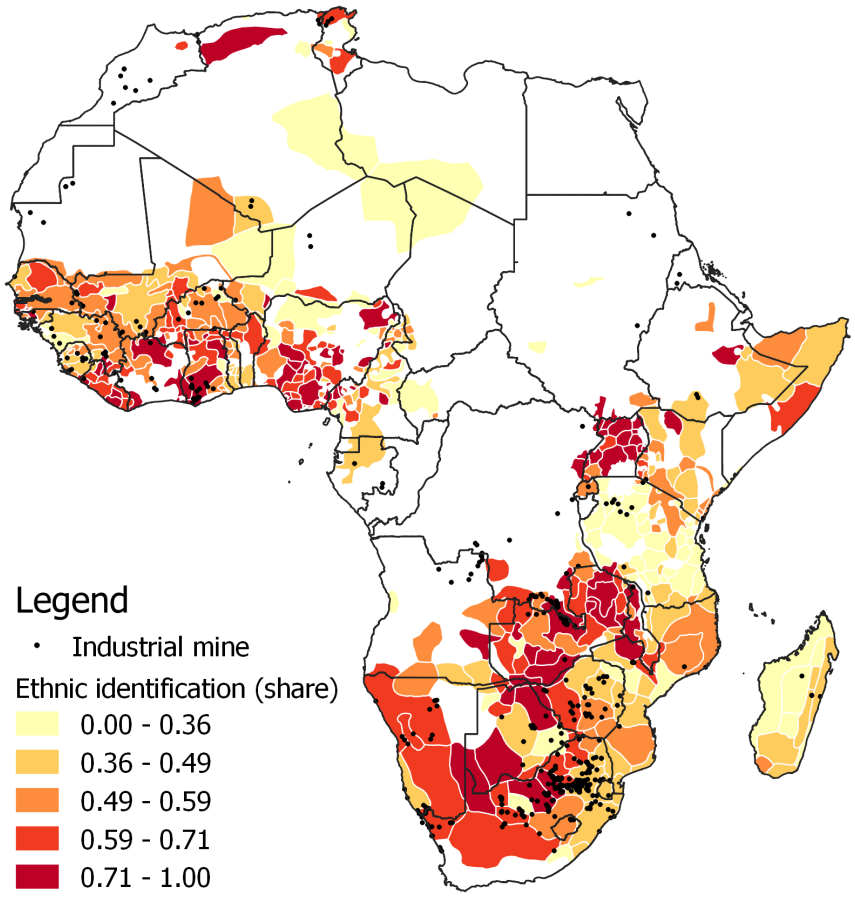
[53] 34% of the Afrobarometer groups are matched to a single one Murdock homeland. For the remaining Afrobarometer groups, the LEDA proposes multiple possible Murdock matches (at the extreme, these go up to 216 possible Murdock matches for the Arabs).

Figure A.3: Identification to ethnic identity and natural resources, Murdock's Cultural goups ethnic homelands



Source: Authors' computations from the Afrobarometer surveys, Murdock and S&P data. The map shows the location of each industrial mine that has been active between the years 2005 and 2015, and the matching of Afrobarometer and Murdock ethnic groups defined at the level of the "Cultural groups". For each group, the map shows the share of members of that group who value their ethnic identity equally or more than their national identity. Each share takes in account the answers of all the members of each ethnic group, independently of the homeland in which these members live.

Figure A.4: Identification to ethnic identity and natural resources, LEDA definition of ethnic homelands



Source: Authors' computations from the Afrobarometer surveys, Murdock and S&P data. The map shows the location of each industrial mine that has been active between the years 2005 and 2015 and the matching of Afrobarometer and Murdock ethnic groups besed on the LEDA (Muller-Crepon et al., 2020). For each ethnic group, the map shows the share of members of that group who value their ethnic identity equally or more than their national identity. Each share takes in account the answers of all the members of each ethnic group, independently of the homeland in which these members live.

A3 Descriptive statistics

Table A.8: Summary statistics for wealth, economic deprivation, conflict and poverty

	Mean	S.D.	1st Quartile	Median	3rd Quartile
Wealth index	-.03	.71	-.25	-.2	.5
Deprivation index	0	.71	-.56	-.08	.52
Economic conditions	2.96	1.05	2	3	4
Group treated unfairly	.7	.96	0	0	1
Battles_{t0}	1.02	3.42	0	.1	.7
$\text{Other conflicts}_{t0}$	3.8	11.89	.3	1.2	3.4
$\text{Nighttime lights}_{t0}$	2.26	3.72	.27	.88	2.23

Source: Authors' computations from the Afrobarometer surveys, Armed Conflict Location Events Data (ACLED) and the National Oceanic and Atmospheric Administration.

A4 Baseline results: full results table

Table A.9: Natural resources extraction and ethnic identification

	(1)	(2)	(3)	(4)	(5)
Dep. Var.	Salience of ethnic group identification				
Version	Categorical	Dummy	Dummy (alt.)	Categorical	Categorical
Sample	Full	Full	Full	Full	Outside homeland
# mines: ethnic homeland	0.018^{a}	0.006^{a}	0.004^{a}		0.016^{a}
	(0.003)	(0.002)	(0.001)		(0.003)
# mines: ethnic homeland, in country				0.017^{a}	
				(0.005)	
# mines: ethnic homeland, abroad				0.018^{a}	
				(0.005)	
Residence in homeland	0.044^{a}	0.014^{a}	0.012^{a}	0.044^{a}	0.000
	(0.012)	(0.005)	(0.003)	(0.012)	(.)
Residence in rural area	0.032^{a}	0.009^{c}	0.009^{a}	0.032^{a}	0.032^{c}
	(0.012)	(0.005)	(0.003)	(0.012)	(0.017)
Female respondent	0.051^{a}	0.019^{a}	0.011^{a}	0.051^{a}	0.048^{a}
	(0.010)	(0.004)	(0.003)	(0.010)	(0.011)
Age	-0.004^{a}	-0.001	-0.002^{a}	-0.004^{a}	-0.007^{a}
	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)
Age squared	0.000^{b}	0.000	0.000^{a}	0.000^{b}	0.000^{a}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
No formal schooling	0.177^{a}	0.050^{b}	0.075^{a}	0.177^{a}	0.092
	(0.053)	(0.023)	(0.016)	(0.053)	(0.064)
Informal schooling only	0.081	0.026	0.046^{a}	0.081	0.014
	(0.057)	(0.025)	(0.016)	(0.057)	(0.070)
Some primary schooling	0.100^{c}	0.033	0.042^{a}	0.100^{c}	0.011
	(0.053)	(0.023)	(0.015)	(0.053)	(0.066)
Primary school completed	0.048	0.021	0.025	0.048	-0.022
	(0.052)	(0.022)	(0.015)	(0.052)	(0.067)
Some secondary school / high school	0.013	0.014	0.005	0.013	-0.061
	(0.052)	(0.023)	(0.016)	(0.053)	(0.067)
Secondary school / high school completed	0.002	0.012	-0.006	0.002	-0.062
	(0.053)	(0.023)	(0.015)	(0.053)	(0.068)
Post-secondary qualifications, other than university	-0.005	0.011	-0.010	-0.005	-0.059
	(0.054)	(0.023)	(0.016)	(0.054)	(0.070)
Some university	-0.028	-0.010	-0.014	-0.028	-0.055
	(0.056)	(0.024)	(0.017)	(0.056)	(0.071)
University completed	0.009	0.001	-0.001	0.009	-0.036
	(0.056)	(0.024)	(0.017)	(0.056)	(0.079)
Inactive	0.050^{a}	0.020^{a}	0.010^{a}	0.050^{a}	0.051^{a}
	(0.013)	(0.006)	(0.003)	(0.013)	(0.016)
Unemployed	0.062^{a}	0.023^{a}	0.011^{a}	0.062^{a}	0.069^{a}
	(0.013)	(0.006)	(0.003)	(0.013)	(0.016)
Employed part time	0.039^{a}	0.017^{a}	0.006^{c}	0.039^{a}	0.041^{b}
	(0.012)	(0.005)	(0.004)	(0.012)	(0.017)
Fixed effects	Country×Ethnic group, Region×Time				
Individual controls			Yes		
Observations	115105	115105	115105	115105	61790
R^2	0.185	0.192	0.110	0.185	0.180
Sample mean / s.d. dep. var.	2.35/1.19	0.53/0.50	0.13/0.34	2.35/1.19	2.31/1.18

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In column (5), we restrict the sample to individuals living outside their ethnic homeland. The dependent variable is the level of ethnic identification with respect to state identification. In columns (1), (4) and (5), the variable ranges from categories 1 to 5. In column (2) we use as dependent variable a dummy taking the value 1 if the individual identifies at least as much to the ethnic group than the country. In column (3) we use an alternative definition of the binary dependent variable, taking the value 1 if the individual identifies strictly more to the ethnic group than the country. There are 1824 Admin-1 region×year×month fixed effects and 401 country×ethnic group fixed effects. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

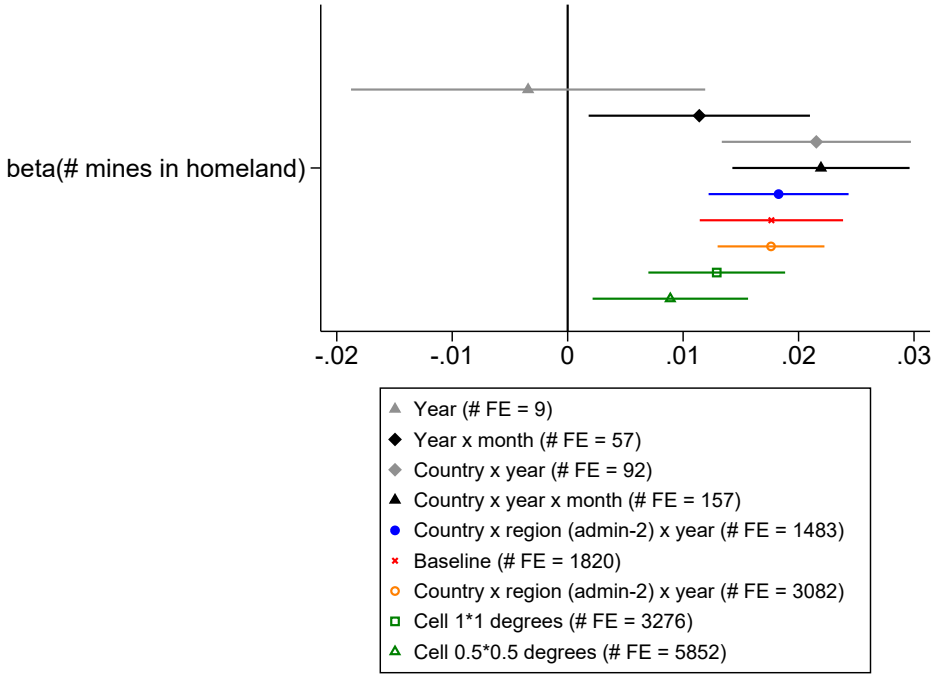
A5 Fixed effects

Table A.10: Fixed effects and variations of # mines

Specification	s.d.	min	max	# FE	Share w/var.> 0 (%)
No controls (intercept)	5.78	-2.06	62.9	-	-
Controls, no FE	5.70	-6.382	64.08	-	-
Ethnic Group FE	1.06	-13.79	12.31	296	16.5
Ethnic Group × Country FE	1.04	-13.74	12.34	396	19.5
Country × Region (admin-1) × Year × Month FE	4.17	-41.56	64.44	1820	51.4
Full specification	0.77	-13.89	12.95	2216	45.7
Country × Region (admin-2) × Year × Month FE	0.76	-14.6	12.37	3082	43.84
Cell 1×1 degrees FE	0.70	-13.49	13.63	3276	52.46
Cell 0.5×0.5 degrees FE	0.67	-12.74	13.35	5852	45.11

This table displays statistics of the # mines in ethnic homeland variable once purged from the contribution of various control variables and sets of fixed effects. We consider as control variables our baseline set of controls: age and its square, gender, rural/urban dummy, education categories and employment status. The "Full specification" is our baseline specification which includes Country × Region (admin-1) × Year × Month FE and Ethnic Group × Country FE.

Figure A.5: Sensitivity of baseline results to alternative sets of fixed effects



Note: this figure reports the estimated coefficients and 95% confidence interval of the variable *# mines in ethnic homeland* based on version of equation (1) where region× time fixed effects are replaced by less demanding sets of fixed effects. In all estimations, ethnic group × country fixed effects are included to ensure that the coefficient is identified across ethnic groups (this implies that country fixed effects are *de facto* included). Two conclusions emerge. First, it is important to control for time-varying local shocks. Controlling for year dummies only leads to insignificant results, and a coefficient close to zero. This makes sense as it is likely that both natural resource exploitation and ethnic feelings have a strong time-varying country-specific component. Second, as long as we control for time-varying local shocks, the definition of "time" (year or year-month) and "local" (country, admin1 region, admin2 region, or apolitical cells) does not affect much the results.

A6 Time to election and binary identification to ethnic groups

Table A.11: Natural resources extraction and binary ethnic identification during electoral periods

Dep. Var.	(1)	(2)	(3)
	Ethnic group identification Dummy		
# mines: ethnic homeland, in country	0.044^{b}		0.134^{a}
	(0.018)		(0.040)
# mines: ethnic homeland × Dist. elec.	-0.006^{b}		
	(0.002)		
# mines: ethnic homeland × Elec. [0-6months]		0.017^{c}	
		(0.009)	
# mines: ethnic homeland × Elec. [6-12months]		0.008	
		(0.007)	
# mines: ethnic homeland × Elec. [12+ months]		0.006	
		(0.006)	
# mines: ethnic homeland × Dist. elec. (after)			-0.008^{a}
			(0.003)
# mines: ethnic homeland × Dist. elec. (before)			-0.011^{a}
			(0.003)
# mines: ethnic homeland, abroad	0.008^{a}	0.008^{a}	0.007^{a}
	(0.002)	(0.002)	(0.002)
Fixed effects	Country×Ethnic group, Region×Time		
Individual controls	Yes		
Observations	109122		
R^2	0.193		
Sample mean / sd dep. var.	0.54/0.50		

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. The dependent variable is a dummy taking the value 1 if the individual identifies at least as much to the ethnic group than the country. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. There are 1767 Admin-1 region×year×month fixed effects and 387 country×ethnic group fixed effects. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

A7 Production value (mineral prices)

Our focus on the number of active mines reflects our interest in how the very existence of these mines affect individual perceptions, potentially independently of these mines' economic reality. Alternatively, we can try to approximate how the real economic value of the mines affect individuals perceptions. This measure is more distant from our research question, but it has the interest of being exogenous as all the identification then relies on yearly variations in world commodity prices which are arguably exogenous to local ethnic identification, especially given that the countries in our sample are typically small producers at the world level.

For each mineral produced in the ethnic homeland of a respondent, the average pre-sample production (2000-2004, or 2004) evaluated yearly at the current real world prices is summed across minerals to get the production value in the ethnic homeland. Formally, the production value in the ethnic homeland h at year t is given by: $Y_{e,t} = \sum_{k=1}^{N} Q_{e,k,t_0} \times P_{k,t}$, where Q_{e,k,t_0} is the production of mineral k in the ethnic homeland e and $P_{k,t}$ is the real world price of the mineral k during year t. As both quantities and prices of different minerals are measured in different units (e.g. metric tons, kilograms, ounces), all prices and quantities are harmonized beforehand (the chosen unit is metric tons, but it does not matter in the outcome variable $Y_{e,t}$).

The outcome variable, $Y_{e,t}$, reflects initial production values evaluated at current prices – it is therefore not a price index. Computing a Laspeyres-like price index would essentially imply multiplying prices by production shares instead of levels. We believe that considering variations in production (solely driven by exogenous prices changes) rather than a price index is preferable, because a price index would mute differences in the size of production. Put differently, we would implicitly expect a similar effect of variations in mineral prices in a homeland hosting, say, one mine producing one kilo of a given mineral, and in a homeland hosting ten mines each producing a thousand ton of the same mineral. This seems like a strong assumption. If our geographical units (ethnic homelands) were very small, this would probably not make a large difference. But given that we are aggregating productions over potentially large geographical units, and units of heterogeneous sizes, we believe that the scale of mining production is important to take into account.

Table A.12: Natural resources and ethnic identification: production value

	(1)	(2)	(3)	(4)
	Salience of ethnic group identification			
Dep Var. Def.	Cat.	Dummy	Cat.	Dummy
Production value: ethnic homeland	0.005	0.002		
ref years: 2000-2004	(0.003)	(0.001)		
Production value: ethnic homeland			0.005	0.002
ref years: 2004			(0.003)	(0.001)
Residence in homeland	0.046^a	0.015^a	0.047^a	0.015^a
	(0.012)	(0.005)	(0.012)	(0.005)
Fixed effects	Country×Ethnic group, Region×Time			
Individual controls	Yes			
Observations	114287	114287	114367	114367
R^2	0.184	0.190	0.184	0.190
Sample mean / sd dep. var.	2.35/1.19	0.54/0.50	2.35/1.19	0.54/0.50

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In columns (1) and (3), the dependent variable is the level of ethnic identification with respect to state identification, ranging from 1 to 5. In column (2) and (4) we use as dependent variable a dummy taking the value 1 if the individual identifies at least as much to the ethnic group than the country. Production value: ethnic homeland: is the price of each mineral weighted by their volume of production at the beginning of the period. Beginning of the period means between the years 2000 and 2004 in columns (1) and (2), or the year 2004 in columns (3) and (4). Individual controls include age and its square, gender, rural/urban dummy, education categories and employment status. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

A8 Weather Shocks

In this section, we study how other ethnic homeland-specific and time-varying shocks – namely weather shocks – affect the impact of mineral exploitation on the strength of ethnic identification. We follow the literature reviewed in Dell et al. (2014) and use either the levels of rainfalls and temperatures or anomalies. More precisely, we compute the average level of each variable in each homeland and year for the period 2000-2015. We first use the levels directly in our estimation and interpret them as deviations from means given our set of fixed effects (like Dell et al., 2012). Alternatively, we use weather anomalies, defined as the variation from the yearly level variable from its long term mean (like Barrios et al., 2010). Table A.13 displays the results: identification to one's ethnic group does not react to any of the two alternative definitions of weather shocks.

Furthermore, we confirm in Table A.14 that the weather shocks taking place in an ethnic homeland correspond to a significant economic shock for households from that ethnic group. The household wealth index varies positively and significantly with rainfall shocks (columns 1 and 3), though not with temperature shocks (columns 4 and 5). When we introduce jointly the rainfall and temperature shocks the results for levels are slightly less precise than results for anomalies in rainfall (p-value at 0.11 in column 4, below 10% in column 6). This finding of a positive and significant relation between rainfalls and wealth aligns with that of Dell et al. (2012) and Barrios et al. (2010), among many others.[54] This finding also alleviates concerns that our wealth proxy may be imprecise and thus strengthens our interpretation of the (lack of) effect of mining on wealth.

Table A.13: Weather shocks and ethnic identification

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Salience of ethnic group identification					
Definition weather variable	Annual averages			Anomalies		
# mines: ethnic homeland	0.018^a	0.018^a	0.018^a	0.018^a	0.018^a	0.018^a
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Rainfall: ethnic homeland	-0.025		-0.027	-0.008		-0.008
	(0.080)		(0.082)	(0.011)		(0.011)
Temperature: ethnic homeland		0.019	0.021		0.007	0.008
		(0.057)	(0.059)		(0.018)	(0.018)
Fixed effects	Country×Ethnic group, Region×Time					
Individual controls	Yes					
Observations	115105	115105	115105	115105	115105	115105
R^2	0.185	0.185	0.185	0.185	0.185	0.185
Sample mean / sd dep. var.	2.35 /1.19					

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. The dependent variable is the level of ethnic identification with respect to state identification in all columns, ranging from categories 1 to 5. In columns 4 to 6, the weather variables capture anomalies, defined for each variable as the deviations from its mean in each homeland, divided by its standard deviation. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

[54] Dell et al. (2012) report a positive but imprecise relation between rainfall and per capita income in poor countries, and a negative relation between temperatures and this income. Barrios et al. (2010) document that higher rainfall is associated with faster growth in these Sub-Saharan African countries.

Table A.14: Weather shocks and household wealth

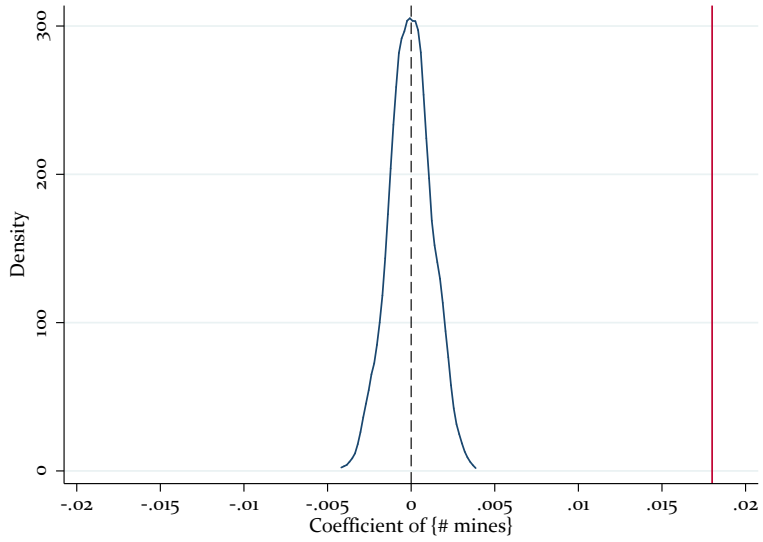
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Wealth index					
Shock definition	Annual averages			Anomalies		
# mines: ethnic homeland	0.006	0.006	0.006	0.006	0.006	0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Rainfall: ethnic homeland	0.071^{c}		0.068	0.011^{b}		0.011^{c}
	(0.041)		(0.042)	(0.006)		(0.006)
Temperature: ethnic homeland		0.028	0.023		0.011	0.010
		(0.032)	(0.032)		(0.009)	(0.009)
Fixed effects	Country×Ethnic group, Region×Time					
Individual controls	Yes					
Observations	116448	116448	116448	116448	116448	116448
R^2	0.322	0.322	0.322	0.322	0.322	0.322
Sample mean / sd dep. var.	-0.028/0.71					

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. The dependent variable is an index of wealth based on ownership of radio, television and vehicles. In columns 4 to 6, weather variables capture anomalies, defined for each variable as the deviations from its mean in each homeland, divided by its standard deviation. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

A9 Random sampling

To further rule out the possibility of false positive in the main results, we run a placebo test as follows: for each year in the sample, we randomly assign the number of mines across ethnic homelands, and re-estimate specification (1) of Table 2 with this random variable. We repeat this Monte Carlo procedure in 1,000 draws. Figure A.6 plots the sampling distribution of the obtained coefficient, compared to our baseline coefficient drawn as a red vertical axis (0.018). Reassuringly, the Monte Carlo coefficients are distributed far from their baseline estimates.

Figure A.6: Monte Carlo Sampling Distribution of (# mines: ethnic homeland)



Note: We draw randomly 1,000 times a number of mines for each homeland, by year, and we estimate specification (1) of Table 2 with this random variable. The figure plots the coefficients obtained, compared to our baseline coefficient drawn as a red vertical axis (0.018).

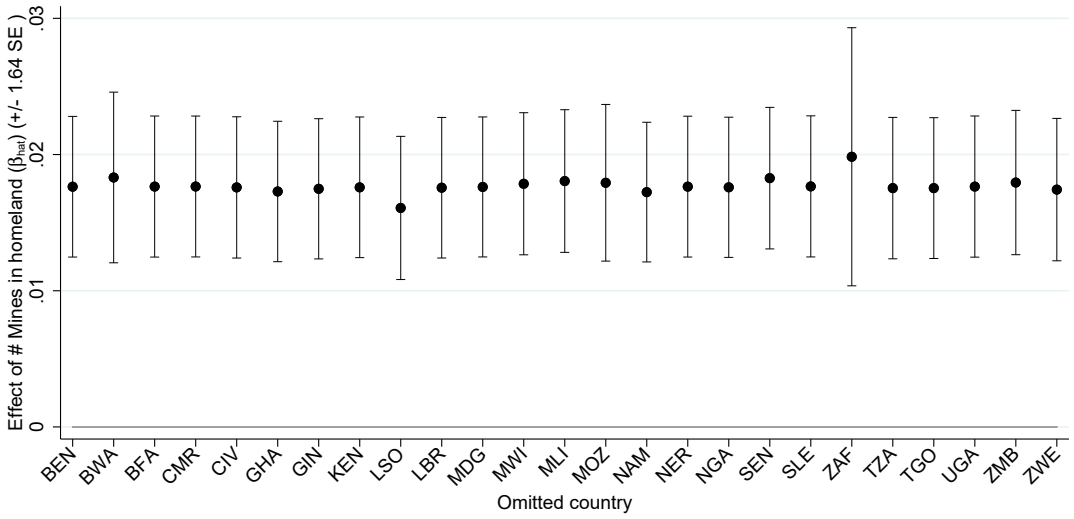
A10 Sensitivity to specific surveys, countries or groups

Table A.15: Natural resources extraction and salience of ethnic identification: dropping survey waves one by one

	(1)	(2)	(3)	(4)
Dep. Var.	Salience of ethnic group identification			
Survey Wave Omitted	3	4	5	6
# mines: ethnic homeland	0.022^a	0.018^a	0.017^a	0.012^b
	(0.005)	(0.004)	(0.004)	(0.006)
Residence in homeland	0.039^a	0.041^a	0.051^a	0.047^a
	(0.014)	(0.013)	(0.015)	(0.012)
Fixed effects	Country×Ethnic group, Region×Time			
Individual controls	Yes			
Observations	94776	92308	78903	79322
R^2	0.174	0.188	0.189	0.194
Sample mean / sd dep. var.	2.3/1.18	2.32/1.18	2.38/1.2	2.4/1.19

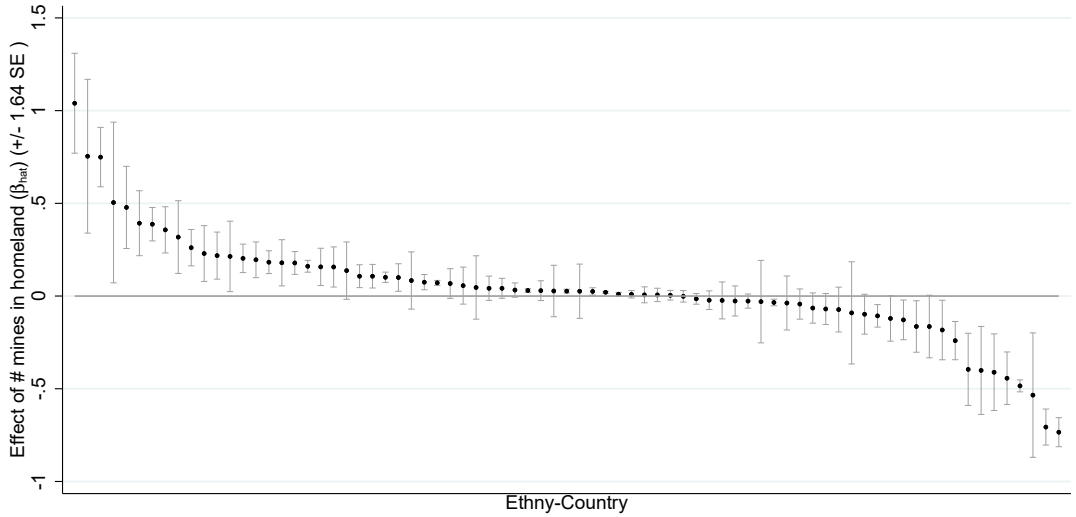
OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. The dependent variable is the level of ethnic identification with respect to state identification in all columns, ranging from categories 1 to 5. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

Figure A.7: Dropping countries one by one



This figure reports our the effect of the number of mines in the ethnic homeland of the respondent on the strength of ethnic identification, estimated using equation (1). Coefficients and 90% confidence interval depicted. Each coefficient is obtained by dropping a specific country from the estimation.

Figure A.8: Baseline results by ethnic group



This figure reports the effect of the number of mines in the ethnic homeland of the respondent on the strength of ethnic identification, estimated using equation (1), except that we allow the coefficient to vary by ethnic group-country by including a set of interaction term between the mining variable and each dummies for each ethnic group-country. Coefficients and 90% confidence interval depicted. The groups with estimated coefficients above 0.3 span over 14 of the 25 countries of the sample. These groups are the CHEWA (whose homeland in Mozambique host a resource, and are present also in Zambia, Malawi and Zimbabwe); the BORAN (from Kenya), the VAI (resource rich in Liberia, also present in Sierra Leone), the GUSII (resource rich in Kenya and Tanzania), the SWAFA (resource rich in Tanzania), the THONGA (resource rich in South Africa and also present in Mozambique, Zambia and Zimbabwe), the TUMBUKA (resource rich in Malawi, and also present in Tanzanie and Zambia), the SOTHO (resource rich in Lesotho and South Africa, and also present in Zimbabwe) and the MOSSI (resource rich in Burkina Faso and also present in Ghana). We note that two of these nine groups having the strongest coefficients are the CHEWA and the TUMBUKA, who are the two groups studied in the influential work of Posner (2004) on the social construction of ethnic identities. Posner (2004) emphasizes how these groups numeric sizes resulted in them being allies in Zambia and opponents in Malawi. Our estimates take the setting emphasized in Posner (2004) as given – as the first of the Afrobarometer round that we use was collected in 2004, and our country × ethnic group level fixed effects absorb the difference in group sizes – and highlights how resources exploitation shape the contemporaneous salience of these ethnic identities.

Table A.16: Outliers Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. var	Salience of ethnic identification							
Sample	Full	Full	Excl. 3-σ outliers	Excl. 2-σ outliers	Excl. 1-σ outliers	Leverage meas.	Influence meas. (dfbeta)	Influence meas. (Cook's distance)
log(1+# mines: ethnic homeland)	0.058							
	(0.036)							
Log Sine Transf.: # mines: ethnic homeland		0.042						
		(0.029)						
# mines: ethnic homeland			0.018^a	0.018^a	0.025^a	0.017^a	0.018^a	0.019^a
			(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
Fixed effects	Country×Ethnic group, Region×Time							
Individual controls	Yes							
Observations	115105	115105	114888	110673	77319	75601	115105	109665
R^2	0.185	0.185	0.194	0.258	0.684	0.164	0.185	0.240

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. The dependent variable is the level of ethnic identification with respect to state identification. In columns (3) to (5), observations that are 3, 2 and 1 standard deviation away from the residual mean are excluded. In column (6), the sample is restricted to observations with high leverage, i.e. when the individual leverage is superior to $2k = N$ (with k being the number of predictors and N the number of observations). In columns (7) and (8), we exclude observations that shift the estimate at least to one standard error (dfbeta) (column 7) and observations that shift the estimate at least to $4 = N$ (Cook's distance), respectively.

A11 Conley standard errors

Table A.17: Natural resources extraction and ethnic identification

	(1)	(2)	(3)	(4)
Dep. Var.	Salience of ethnic group identification			
Version	Categorical	Dummy	Categorical	Categorical
Sample	Full	Full	Full	Outside homeland
# mines: ethnic homeland	0.018	0.006		0.016
Ethnic group level: baseline	$(0.003)^a$	$(0.002)^a$		$(0.003)^a$
Distance: 100	$(0.006)^a$	$(0.003)^c$		$(0.007)^b$
Distance: 250	$(0.006)^a$	$(0.003)^c$		$(0.007)^b$
Distance: 500	$(0.004)^a$	$(0.002)^b$		$(0.005)^a$
Distance: 750	$(0.004)^a$	$(0.002)^b$		$(0.005)^a$
Distance: 1000	$(0.004)^a$	$(0.002)^b$		$(0.005)^a$
# mines: ethnic homeland, in country			0.017	
Ethnic group level: baseline			$(0.005)^a$	
Distance: 100			$(0.007)^b$	
Distance: 250			$(0.006)^a$	
Distance: 500			$(0.005)^a$	
Distance: 750			$(0.005)^a$	
Distance: 1000			$(0.005)^a$	
# mines: ethnic homeland, abroad			0.018	
Ethnic group level: baseline			$(0.005)^a$	
Distance: 100			(0.013)	
Distance: 250			(0.012)	
Distance: 500			$(0.010)^c$	
Distance: 750			$(0.009)^b$	
Distance: 1000			$(0.009)^b$	
Residence in homeland	0.044	0.014	0.044	
Ethnic group level: baseline	$(0.012)^a$	$(0.005)^a$	$(0.012)^a$	
Distance: 100	$(0.011)^a$	$(0.004)^a$	$(0.012)^a$	
Distance: 250	$(0.012)^a$	$(0.005)^a$	$(0.012)^a$	
Distance: 500	$(0.013)^a$	$(0.005)^a$	$(0.013)^a$	
Distance: 750	$(0.012)^a$	$(0.005)^a$	$(0.013)^a$	
Distance: 1000	$(0.013)^a$	$(0.005)^a$	$(0.013)^a$	
Fixed effects	Country×Ethnic group, Region×Time			
Individual controls	Yes			
Observations	115105	115105	115105	61790
R^2	0.185	0.192	0.185	0.180

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are estimated allowing for infinite serial correlation and spatial correlation for different thresholds (errors estimated for each threshold appear in a different line). In columns (4), we restrict the sample to individual living outside their ethnic homeland. The dependent variable is the level of ethnic identification with respect to state identification in all columns but the last one, ranging from categories 1 to 5. In column (2) we use as dependent variable a dummy taking the value 1 if the individual identifies at least as much to the ethnic group than the country. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. There is 1824 country×region×year×month fixed effect and 401 country×ethnic group fixed effects. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

A12 Alternative group matching

Table A.18: Natural resources extraction and ethnic identity (baseline)

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Salience of ethnic group identification (Categorical)					
Ethnic Homeland Def.	Cultural		LEDA		EPR	
Sample	Full	Outside	Full	Outside	Full	Outside
# mines: ethnic homeland	0.005[a]	0.007[a]	0.018[a]	0.017[a]	0.007[c]	0.009[a]
	(0.001)	(0.002)	(0.004)	(0.004)	(0.003)	(0.003)
Residence in homeland	0.030[b]		0.049[a]		0.057[a]	
	(0.014)		(0.013)		(0.018)	
Fixed effects	Country×Ethnic group, Region×Time					
Individual controls	Yes					
Observations	115101	38446	116117	74591	84841	35996
R^2	0.181	0.182	0.186	0.181	0.177	0.167
Dep. var sample mean/s.d.	2.35/1.19	2.32/1.18	2.34/1.19	2.33/1.18	2.36/1.2	2.31/1.18

OLS estimations. [c] significant at 10%; [b] at 5%; [a] at 1%. Standard errors are clustered at the ethnic group level. In columns (2) (4) and (6), we restrict the sample to individual living outside their ethnic homeland. The dependent variable is the level of ethnic identification with respect to state identification, ranging from categories 1 to 5. Columns (1) and (2) use the version of Murdock ethnic homelands based on "cultural groups". Column (3) and (4) use the matching based on LEDA. Columns (5) and (6) uses ethnic groups homelands from EPR. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

Table A.19: Natural resources extraction and ethnic identity (dummy)

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Salience of ethnic group identification (Dummy, base)					
Ethnic Homeland Def.	Cultural		LEDA		EPR	
Sample	Full	Outside	Full	Outside	Full	Outside
# mines: ethnic homeland	0.003[a]	0.004[a]	0.005[a]	0.004	0.002	0.003[b]
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Residence in homeland	0.009		0.016[a]		0.016[b]	
	(0.006)		(0.005)		(0.006)	
Fixed effects	Country×Ethnic group, Region×Time					
Individual controls	Yes					
Observations	115101	38446	116117	74591	84841	35996
R^2	0.181	0.182	0.186	0.181	0.177	0.167
Dep. var sample mean/s.d.	2.35/1.19	2.32/1.18	2.34/1.19	2.33/1.18	2.36/1.2	2.31/1.18

OLS estimations. [c] significant at 10%; [b] at 5%; [a] at 1%. Standard errors are clustered at the ethnic group level. In columns (2) (4) and (6), we restrict the sample to individual living outside their ethnic homeland. The dependent variable a dummy taking the value 1 if the individual identifies at least as much to the ethnic group than the country. Columns (1) and (2) use the version of Murdock ethnic homelands based on "cultural groups". Column (3) and (4) use the matching based on LEDA. Columns (5) and (6) uses ethnic groups homelands from EPR. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

A13 Switches in ethnic identity

Table A.20: Natural resources extraction and switches in ethnic group identity

	(1)	(2)	(3)	(4)	(5)
Dep. Var.	Share ethnic group				Group dummy
Ethnic Homeland Def.	Baseline	Cultural	LEDA	EPR	Baseline
# mines: ethnic homeland	-0.000	0.001	-0.002	0.001	-0.001[c]
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Country × year FE	Yes	Yes	Yes	Yes	No
Ethnic group × Country	Yes	Yes	Yes	Yes	No
Country×region×year×month FE	No	No	No	No	Yes
Country×ethnic group FE	No	No	No	No	Yes
Additional controls		Yes			
Observations	1175	542	1083	564	1784874
R^2	0.982	0.989	0.906	0.747	0.302
Sample mean	0.066	0.144	0.061	0.134	0.066

OLS estimations. [c] significant at 10%; [b] at 5%; [a] at 1%. Standard errors are clustered at the ethnic group level. In columns (1) to (4) the dependent variable is the share of Afrobarometer respondents who declare they belong to that ethnic group, computed by country and wave of survey. Columns (1) to (3) use different versions of Murdock ethnic homelands, as defined in the text. Column (4) use ethnic groups homelands from EPR. In column (5) the level of observation is the individual-ethnic group. The dependent variable is a dummy taking the value 1 if the individual declares belonging to that ethnic group. We consider all combinations respondent × ethnic group observed in the country. Controls include age and its square, gender, rural/urban dummy, education categories and employment status.

A14 Countries background

A14.1 Zambia

Similar to many countries in Sub-Saharan Africa, Zambia's territory overlaps with different historical ethnic homelands, and hosts significant natural resources. The country hosts a variety of ethnic groups, the main three being the Bemba (21% of the population), the Tonga (13.6%) and the Chewa (7.4%) (CIA Factbook). The political rise of Michael Sata was heavily linked to the Bemba group. Commercial copper mining in Zambia started in 1928 and the revenues of copper exploitation have shaped the economy of the country (Lungu, 2008). The country is rich in other minerals, but copper is the main one. In 2019, copper represented over 73% of the country's 10.1 billion dollars value of exports (53% for raw copper, and 19% refined copper).[55] Zambia's copper producing region – in the so-called "Copperbelt" – illustrates the nexus between the effect of natural resource exploitation, political discourse and ethnic identities, as well as how these effects may travel across space through co-ethnic linkages. Though the Copperbelt mining towns are ethnically diverse, the Bemba group represents a majority among Copperbelt workers (Cheeseman and Larmer, 2015). For decades, parties like the United National Independence Party (UNIP) – since the 1960s – or Michael Sata's Patriotic Front (PF) in the 2000s have practiced a form of "ethnopopulism" in the mining area, playing with grievances and expectations of redistribution of mining rents. This ethnic dimension of the movement appears to be articulated with an economic dimension as in "Zambia, widespread discontent arose from the fact that the high value of Copperbelt minerals resulted in little material benefit for Copperbelt residents. This discontent was articulated by Michael Sata's Patriotic Front Party" (Larmer, 2016). These made identities more salient, especially among the Bemba group: to cite Cheeseman and Larmer (2015, p.35) "mineworkers and Copperbelt residents of all ethnic groups expected more of the wealth generated by mining to be spent on raising wages and working conditions, whilst the linkage with, and dependence on, urban areas made rural Bembas more responsive to populist messages than other rural Zambians". Hence, not only mining-related grievances were exacerbated, but these likely reinforced the identity of a specific ethnic group, both in the copper-producing region and outside the region.

A14.2 Indonesia

Indonesia hosts a variety of ethnic groups, the main one being Javanese, who represent 40.1% of the population (CIA Factbook). The two main other groups are the Sundanese (15.5%) and the Malay (3.7%). The Acehnese, the main base of the Free Aceh movement studied by Aspinall (2007), are only the twelfth biggest group, representing 1.4% of the country's population. Indonesia hosts natural resources in many parts of the archipelago. In 2019, oil and gas represented about 13% of the country's 165 billion dollars value of exports (7% for refined petroleum, and 3% crude petroleum).[56] Gas reserves were discovered in Aceh in 1971, which became one of the major producing provinces. The conflict in Aceh started in 1988 and appears to be heavily linked to resources exploitation, though indirectly. Aspinall (2007) argues that "the evolving framework of Acehnese identity that provided a prism through which natural resource exploitation was

[55] Source: https://oec.world/en/profile/country/zmb
[56] Source: https://oec.world/en/profile/country/idn

interpreted in grievance terms." (Aspinall, 2007, p.957) The claims of "unjust exploitation" of natural resources, pushed by "ethnic political entrepreneurs", reinforced the discourse of deprivation that pre-existed in Aceh. This echoes the results that we find in the case of minerals in Sub-Saharan Africa: the impact on ethnic identities appears stronger when mines open in regions with a pre-existing history of conflict, or where pre-existing levels of ethnic identification are high.

A14.3 Tanzania

Tanzania is considered as a success story of national integration, creating a national identity above the pre-existing local cleavages (Green, 2013; Miguel, 2004). The CIA Factbook, for example, does not list the same level of details of ethnic groups as for other countries. Rather, they indicate that 99% of the country's population is mainland Africans, of which 95% are Bantu consisting of more than 130 tribes. This description is in sharp contrast with Ndulu et al. (2019), noting that "With an ELF index of 0.93 and the politically relevant ethnic groups value of 0.59, Tanzania is clearly one of the most ethnically fractionalised countries in Africa, a home to many ethnic groups with small or relatively equal percentages of the population." Over the last decades, Tanzanian politics have started to be affected by repeated major oil and gas discoveries. The production was quasi nonexistent till the 2000, when it started increasing with a boom in the 2010s. Recent tensions in the country have been attributed to grievances around the extractive activities. Must and Rustad (2016, 2019) show how unmet expectations of increased revenues and better public services triggered social unrest in the southern Mtwara region, which hosts a marginalized population with a strong group identity. According to Must and Rustad, key elements leading to this situation were the widespread perceptions of inequality, injustice and unequal treatment from the government. These feelings of 'relative deprivation' generated grievances and triggered violent social unrest. Opposition parties played their part by organizing large community meetings, "capitalizing on the sentiments among the locals" (Must and Rustad, 2016, p.3). This case study fits quite well with our findings: though economic conditions did not appear to have actually worsened in Mtwara, expectations of local redistribution and strong local collective identity, partly influenced by political leaders, resulted in violent conflict. Grievances and identities were used by elites as mobilization tools.

A15 Wealth and deprivation: full results

Table A.21: Natural resources extraction, wealth, and deprivation

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Wealth index		Deprivation index		Economic conditions	
Sample	Full	Outside homeland	Full	Outside homeland	Full	Outside homeland
# mines: ethnic homeland	0.006	0.004	0.004^{b}	0.008^{a}	-0.003	-0.008^{c}
	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)
Residence in homeland	-0.028^{a}		0.020^{b}		0.003	
	(0.006)		(0.009)		(0.011)	
Residence in rural area	-0.216^{a}	-0.216^{a}	0.109^{a}	0.109^{a}	-0.036^{a}	-0.041^{a}
	(0.009)	(0.012)	(0.010)	(0.014)	(0.010)	(0.012)
Female	-0.137^{a}	-0.134^{a}	-0.013^{a}	-0.011^{c}	-0.012^{c}	-0.011
	(0.013)	(0.013)	(0.005)	(0.006)	(0.008)	(0.009)
Age	0.028^{a}	0.031^{a}	0.010^{a}	0.011^{a}	-0.013^{a}	-0.014^{a}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Age squared	-0.000^{a}	-0.000^{a}	-0.000^{a}	-0.000^{a}	0.000^{a}	0.000^{a}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
No formal schooling	-0.904^{a}	-0.897^{a}	0.496^{a}	0.488^{a}	-0.295^{a}	-0.333^{a}
	(0.038)	(0.041)	(0.028)	(0.036)	(0.053)	(0.064)
Informal schooling only	-0.756^{a}	-0.746^{a}	0.452^{a}	0.444^{a}	-0.261^{a}	-0.300^{a}
	(0.038)	(0.044)	(0.034)	(0.043)	(0.051)	(0.064)
Some primary schooling	-0.739^{a}	-0.741^{a}	0.410^{a}	0.394^{a}	-0.250^{a}	-0.275^{a}
	(0.039)	(0.041)	(0.028)	(0.035)	(0.050)	(0.061)
Primary school completed	-0.623^{a}	-0.624^{a}	0.345^{a}	0.344^{a}	-0.219^{a}	-0.254^{a}
	(0.040)	(0.043)	(0.027)	(0.035)	(0.050)	(0.062)
Some secondary school / high school	-0.543^{a}	-0.542^{a}	0.291^{a}	0.279^{a}	-0.174^{a}	-0.207^{a}
	(0.038)	(0.040)	(0.026)	(0.033)	(0.049)	(0.060)
Secondary school / high school completed	-0.428^{a}	-0.423^{a}	0.221^{a}	0.210^{a}	-0.122^{b}	-0.153^{b}
	(0.040)	(0.041)	(0.026)	(0.035)	(0.052)	(0.063)
Post-secondary qualifications	-0.244^{a}	-0.236^{a}	0.134^{a}	0.131^{a}	-0.080	-0.112^{c}
	(0.038)	(0.042)	(0.027)	(0.034)	(0.055)	(0.062)
Some university	-0.309^{a}	-0.311^{a}	0.107^{a}	0.121^{a}	-0.088	-0.127^{c}
	(0.035)	(0.041)	(0.029)	(0.038)	(0.058)	(0.068)
University completed	-0.052	-0.043	0.028	0.015	-0.007	-0.019
	(0.033)	(0.040)	(0.029)	(0.037)	(0.052)	(0.066)
Inactive	-0.209^{a}	-0.213^{a}	0.086^{a}	0.077^{a}	-0.062^{a}	-0.072^{a}
	(0.011)	(0.011)	(0.008)	(0.010)	(0.015)	(0.016)
Unemployed	-0.228^{a}	-0.232^{a}	0.173^{a}	0.172^{a}	-0.147^{a}	-0.138^{a}
	(0.012)	(0.011)	(0.011)	(0.012)	(0.015)	(0.016)
Working part time	-0.100^{a}	-0.110^{a}	0.101^{a}	0.105^{a}	-0.064^{a}	-0.077^{a}
	(0.011)	(0.010)	(0.011)	(0.013)	(0.019)	(0.019)
Fixed effects	Country×Ethnic group, Region×Time					
Individual controls	Yes					
Observations	117775	63815	117788	63821	80667	43438
R^2	0.325	0.337	0.263	0.283	0.154	0.154
Sample mean / sd dep. var.	-0.02/0.71	0.11/0.72	-0.017/0.71	-0.026/0.72	2.95/1.05	2.99/1.035

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In columns (2) (4) and (6), we restrict the sample to individuals living outside their ethnic homeland. The dependent variable in columns (1) and (2) is an index of wealth based on ownership of radio, television and vehicles. In columns (3) and (4), it is the perceived deprivation index, based on whether the respondent declares having gone without the following items over the last year: food, income, water, medicine, fuel. In columns (5) and (6), the dependent variable is a categorical variable measuring the respondent's perception of his/her own living conditions compared to twelve months before. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

A16 The role of mines' characteristics

Following Pelzl and Poelhekke (2021), we classify as labor intensive the mines which use (at least partly) underground extractions method; all other methods of extraction, mostly open pit mines, placer mines and tailings are classified as capital intensive. To investigate the role of ownership structure, we count separately the number of mines in each homeland that are owned by either foreign or domestic, entities.[57]

The results, shown in Table A.22, suggest that domestic mines trigger stronger ethnic feelings and feelings of deprivation. However, the coefficients of the foreign and domestic mines variables are quite noisy and are generally not statistically different from each other, except in the case of economic deprivation. Consistent with Pelzl and Poelhekke (2021), only relatively capital intensive mines are associated with improvements in the wealth index, though again the coefficients are too noisy to conclude that their impact is statistically different from that of labor intensive mines. In addition, we do not detect any substantial difference in terms of ethnic identification or feelings of deprivation. Overall we interpret these findings as suggestive that the existence of local economic effects – and hence the payoffs associated with belonging to the local ethnic group – do not appear to be prime determinants of ethnic identification.

Table A.22: The role of mines' characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ethnic identification		Wealth index		Deprivation index		Economic conditions	
# mines in homeland : domestic	0.023^{a}		0.006		0.010^{a}		-0.007	
	(0.006)		(0.004)		(0.003)		(0.007)	
# mines in homeland : foreign	0.012^{c}		0.007		-0.002		-0.001	
	(0.006)		(0.005)		(0.003)		(0.005)	
# mines in homeland : L-intens		0.020^{b}		-0.000		0.006^{c}		-0.019
		(0.010)		(0.007)		(0.004)		(0.016)
# mines in homeland : others		0.017^{a}		0.009^{b}		0.003		0.000
		(0.006)		(0.004)		(0.002)		(0.004)
Residence in homeland	0.044^{a}	0.044^{a}	-0.028^{a}	-0.028^{a}	0.020^{b}	0.020^{b}	0.003	0.003
	(0.012)	(0.012)	(0.006)	(0.006)	(0.009)	(0.009)	(0.010)	(0.010)
Diff. in mines coefs.	0.011	0.002	-0.001	-0.009	0.013^{a}	0.003	-0.006	-0.019
	(0.010)	(0.010)	(0.014)	(0.007)	(0.005)	(0.005)	(0.009)	(0.017)
Fixed effects	Country×Ethnic group, Region×Time							
Individual controls	Yes							
Observations	115105	115105	117775	117775	117788	117788	80667	80667
R^2	0.185	0.185	0.325	0.325	0.263	0.263	0.154	0.154
Sample mean / sd dep. var.	2.34/1.19		-0.024/0.71		-0.002/ 0.71		2.96/1.04	

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. The dependent variable in columns (1) and (2) is our baseline categorical measure of ethnic identification. In columns (3) and 4), it is an index of wealth based on ownership of radio, television and vehicles. In columns (5) and (6), it is the perceived deprivation index, based on whether the respondent declares having gone without the following items over the last year: food, income, water, medicine, fuel. Finally, in columns (7) and (8), the dependent variable is a categorical variable measuring the respondent's perception of his/her own living conditions compared to twelve months before. Controls include age and its square, gender, rural/urban dummy, education categories, employment status and a dummy for residence in the ethnic homeland. Domestic mines are mines which are at least partially owned by the state of the country where they are located. Foreign mines are non-domestic mines. L-intensive mines are mines which use at least partly underground extraction method. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

[57] We consider a mine as domestic if it is at least partially owned by the state of the country where it is located.

A17 Local effects: production value

Table A.23: Natural resources extraction: local effects – mining production

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Wealth index		Deprivation index		Economic conditions	
Sample	Full	Residents	Full	Residents	Full	Residents
Prod. value $[0-50]$ outside hom.	0.004^{b}	0.072	0.002	-0.066^{b}	0.003	0.227^{a}
	(0.002)	(0.068)	(0.002)	(0.030)	(0.003)	(0.072)
Prod. value $[51-150]$ outside hom.	-0.000	0.010^{c}	0.001	-0.004	0.002^{c}	0.004
	(0.001)	(0.005)	(0.001)	(0.003)	(0.001)	(0.007)
Prod. value $[0-50]$ in hom.	0.001	0.011^{c}	0.006	0.018^{c}	0.002	-0.008
	(0.007)	(0.006)	(0.009)	(0.010)	(0.004)	(0.011)
Prod. value $[51-150]$ in hom.0	0.004^{c}	0.005	0.004^{b}	0.008^{c}	-0.000	0.002
	(0.002)	(0.006)	(0.002)	(0.004)	(0.004)	(0.014)
Fixed effects	Country×Ethnic group, Country×Time					
Individual controls	Yes					
Diff in hom./outside hom.	0.001	-0.066	0.006	0.096^{a}	-0.003	-0.238^{a}
	(0.009)	(0.069)	(0.010)	(0.036)	(0.007)	(0.077)
Observations	117930	53915	117943	53922	80725	37169
R^2	0.295	0.295	0.198	0.202	0.103	0.124
Sample mean / sd dep. var.	-0.024/0.71	-0.07/0.7	-0.002/0.71	0.03/0.71	2.95/1.04	2.9/1.06

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In columns (2) (4) and (6), we restrict the sample to individual living inside their ethnic homeland. The dependent variable in columns (1) and (2) is an index of wealth based on ownership of radio, television and vehicles. In columns (3) and (4), it is the perceived deprivation index, based on whether the respondent declares having gone without the following items over the last year: food, income, water, medicine, fuel. In columns (5) and (6), the dependent variable is a categorical variable measuring the respondent's perception of his/her own living conditions compared to twelve months before. Prod. value is the total value of production of the mines located within the indicated radius, computed for each mine as the weighted sum of pre-period (2000-2004) volume of production of each minerals multiplied by its real world prices. Controls include age and its square, gender, rural/urban dummy, education categories, employment status and, in columns (1) to (6), a dummy for residence in the ethnic homeland. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.

A18 Excluding EPR Sample

Table A.24: Natural resources extraction and ethnic identification, excluding EPR sample

	(1)	(2)	(3)	(4)
Dep. Var.	Salience of ethnic group identification			
Version	Categorical	Dummy	Categorical	Categorical
Sample	Full	Full	Full	Outside homeland
# mines: ethnic homeland	0.048^a	0.025^a		0.054^a
	(0.016)	(0.007)		(0.019)
# mines: ethnic homeland, in country			0.071^b	
			(0.034)	
# mines: ethnic homeland, abroad			0.038^a	
			(0.013)	
Residence in homeland	-0.012	-0.010	-0.012	
	(0.043)	(0.022)	(0.043)	
Fixed effects	Country×Ethnic group, Region×Time			
Individual controls	Yes			
Observations	34559	34559	34559	21582
R^2	0.203	0.211	0.203	0.170
Sample mean / s.d. dep. var.	2.31/1.16	0.52/0.50	2.31/1.16	2.35/1.13

OLS estimations. ^{c} significant at 10%; ^{b} at 5%; ^{a} at 1%. Standard errors are clustered at the ethnic group level. In column (4), we restrict the sample to individual living outside their ethnic homeland. The dependent variable is the level of ethnic identification with respect to state identification in all columns but one, ranging from categories 1 to 5. In column (2) we use as dependent variable a dummy taking the value 1 if the individual identifies at least as much to the ethnic group than the country. Controls include age and its square, gender, rural/urban dummy, education categories and employment status. There are 1824 Admin-1 region×year×month fixed effects and 401 country×ethnic group fixed effects. Sample mean / sd dep. var. are respectively the sample mean and standard deviation of the dependent variable in the corresponding column.