

DISCUSSION PAPER SERIES

DP15162

**CASH CROPS, PRINT TECHNOLOGIES
AND THE POLITICIZATION OF
ETHNICITY IN AFRICA**

Yannick Pengl, Philip Roessler and Valeria Rueda

ECONOMIC HISTORY



CASH CROPS, PRINT TECHNOLOGIES AND THE POLITICIZATION OF ETHNICITY IN AFRICA

Yannick Pengl, Philip Roessler and Valeria Rueda

Discussion Paper DP15162
Published 13 August 2020
Submitted 12 August 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Economic History

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Yannick Pengl, Philip Roessler and Valeria Rueda

CASH CROPS, PRINT TECHNOLOGIES AND THE POLITICIZATION OF ETHNICITY IN AFRICA

Abstract

What are the origins of the ethnic landscapes in contemporary states? Drawing on a pre-registered research design, we test the impact of dual socioeconomic revolutions that spread across Africa in the 19th and 20th centuries - export agriculture and print technologies. We argue these changes transformed ethnicity via their effects on politicization and boundary-making. Print technologies strengthened imagined communities, leading to more salient yet porous-ethnic identities. Cash crop endowments increased groups' mobilizational potential but with more exclusionary boundaries to control agricultural rents. Using historical data on cash crops and African language publications, we find that groups exposed to these historical forces are more likely to be politically relevant in the post-independence period, and their members report more salient ethnic identities. We observe heterogenous effects on boundary-making as measured by inter-ethnic marriage; relative to cash crops, printing fostered greater openness to assimilate linguistically-related outsiders. Our findings not only illuminate the historical sources of ethnic politicization, but mechanisms shaping boundary formation.

JEL Classification: N47, N57, O13, O43, Z12, Z13

Keywords: Africa, Agriculture, Colonialism, Ethnicity, Missions, Language

Yannick Pengl - yannick.pengl@icr.gess.ethz.ch
ETH Zurich

Philip Roessler - proessler@wm.edu
William and Mary

Valeria Rueda - ruedavaleria@gmail.com
University of Nottingham, School of Economics, Economic and Social History Group, Oxford and CEPR

Cash Crops, Print Technologies and the Politicization of Ethnicity in Africa

Yannick I. Pengl *ETH Zurich*

yannick.pengl@icr.gess.ethz.ch

Philip Roessler *William & Mary*

proessler@wm.edu

Valeria Rueda *University of Nottingham*

valeria.rueda@nottingham.ac.uk

August 12, 2020

Abstract. What are the origins of the ethnic landscapes in contemporary states? Drawing on a pre-registered research design, we test the impact of dual socioeconomic revolutions that spread across Africa from the 19th century onward—export agriculture and print technologies. We argue these changes transformed ethnicity via their effects on politicization and boundary-making. Print technologies strengthened imagined communities, leading to more salient—yet porous—ethnic identities. Cash crop endowments increased groups' mobilizational potential but with more exclusionary boundaries to control agricultural rents. Using historical data on cash crops and African language publications, we find that groups exposed to these historical forces are more likely to be politically relevant after independence, and their members report stronger ethnic identities. We observe heterogeneous effects on boundary-making as measured by inter-ethnic marriage; relative to cash crops, printing fostered greater openness to assimilate linguistically-related outsiders. Our findings illuminate the historical sources of ethnic politicization and mechanisms shaping boundary formation.

This research is supported by the US National Science Foundation (award #1628498) and the Swiss National Science Foundation (grant #P0EZP1_159076). We thank Paola Galano Toro, Vanessa Kellerhals, Benjamin Füglistter, Lukas Dick, Carlos Mairoce, and Julian Seitlinger (ETH Zurich) as well as Layla Abi-Falah, Sidhart Bhushan, Hedda Roberts, Aaron Spitler, and Henry Young (W&M) for excellent research assistance. Earlier versions of our research design were presented at WGAPPE, LSE, March 2018; Annual Meeting of the American Political Science Association, Boston, September 2018; Princeton University Comparative Politics Colloquium, October 2018; Annual Meeting of the AEHN Bologna, October 2018. First results and early paper versions presented at the Annual Meeting of the American Political Science Association, Washington DC, September 2019; Political Economy Workshop, Zurich, October 2019. We are grateful to participants for their suggestions and feedback. Special thanks to Matthew Gichohi, Carl Müller-Crepon, and Lars-Erik Cederman for detailed comments.

Introduction

What are the origins of the ethno-political landscapes that shape contemporary states? A voluminous literature points to the impact ethnicity — social identity based on shared descent and culture (Fearon 2003; Chandra 2006) — has on politics and the allocation of state resources. From taxation and the provision of public services to civil war, ethnicity is found to structure a wide-range of political and economic processes (kasara; Horowitz 1985; Chandra 2004; Habyarimana et al. 2009; Roessler 2016). Less well-understood, however, are the causes of the *politicization of ethnicity*—that is why some ethnic groups, but not others, serve as the bases of political mobilization (Bates 1983; Fearon 2006). This represents a major gap. Politicization is an important phenomenon in its own right. Moreover, without knowing the sources of the ethno-political configurations that shape policy outcomes, the ethnic politics literature is vulnerable to potential selection problems (Birbir et al. 2015; Birbir et al. 2018)

Our paper contributes to closing this gap. We focus on sub-Saharan Africa, a region in which ethnicity has been central to politics, but also in which there is a significant disparity in the number of ethno-linguistic groups and politically-relevant ones.¹ Why do some cultural groups, but not others, serve as bases for political competition? And what accounts for the boundary configurations that shape these groups? We address these two questions, reporting the results of a pre-registered research design.²

We propose that Africa’s ethno-political landscape was shaped by dual socioeconomic revolutions that swept across the continent beginning from the 19th century: the spread of commercial export agriculture and the diffusion of writing and printing technologies.³ We posit that both transformations activated and politicized ethnicity by endowing groups with, on the one hand, greater access to material resources and, on the other hand, improved communication capabilities. These changes had a two-fold effect. They strengthened groups’ mobilizational

¹In our sample of 35 Sub-Saharan African countries, there exist 2303 Ethnologue languages, whereas the Ethnic Power Relations dataset counts anywhere from 170 to 199 groups politically relevant (between 1980 and 2005) (Vogt et al. 2015).

²We pre-registered our research design with Evidence in Governance and Politics (EGAP) on April 24, 2019 after some promising preliminary analyses (geographic correlations between cash crops and Afrobarometer/DHS outcomes as well as between proximity to missionary printing presses and Afrobarometer identity salience) but *before* merging our publications and cash crop data with Ethnologue language categories and group polygons and, via Ethnologue, to ethnic categories in EPR, PREG, Afrobarometer and DHS. An anonymized version of our pre-analysis plan can be found here: [Link to PDF](#).

³Throughout this paper we interchangeably use the terms print technology, publishing, and communication technologies as shorthand to account for the process by which formerly oral languages became written and printed. As we discuss in detail below, this entailed, first, language standardization, then the development of a writing system, and ultimately printing and publishing—the indicator we use to measure this process.

capacity, enhancing their ability to compete in the political arena. They also increased ethnicity's salience through the commercialization of land and the adoption of descent-based property rights regimes, and greater exposure to and consumption of a uniform set of cultural characteristics, texts, and histories.

Whereas we theorized that the diffusion of cash crops and print technologies produced similar effects on ethnic politicization, our expectation was that they induced heterogeneous effects on boundary-making—the social boundaries that regulate group membership (Barth 1969; Wimmer 2013)—due to competing processes of *social closure* (Weber 1978; Parkin 1979) and *imagined communities* (Anderson 2006). Africa's print revolution was catalyzed by Christian missionaries intent on increasing access to translated Bibles to spread Christianity (Ranger 1989; Posner 2003). We hypothesized that the diffusion of print technologies and publishing in vernacular languages contributed to the transmission of shared cultural beliefs and practices, creating a vehicle for social assimilation and leading to more inclusionary ethnic boundaries. In contrast, we advanced that the cash crop revolution led to more exclusionary social identities and less ethnic assimilation. We attribute this to the Weberian process of social closure (Parkin 1979)⁴ in which sons-of-the-soil, responding to growing demand for access to their agricultural-rich homeland, including from migrant farmers and laborers, used ethnic identity as a means to regulate land ownership and control agricultural rents.

To test the effect of these historical processes on ethnic politicization and boundary-making we combine detailed historical data on cash crop production and publications in African languages with contemporary ethnicity data. Our cash crop data is based on a comprehensive historical map on the source locations of exports in late colonial Africa (Hance, Kotschar, and Peterec 1961). It identifies 9,517 geocoded production points valued at \$289,270 per point, covering 9 groups of cash crops, 20 minerals and metals, and forest, animal and manufactured products. We compile a novel dataset of historical African language publications from Rowling and Wilson (1923) and Mann and Sanders (1994). Together, these two bibliographic sources cover approximately 10,000 titles in 370 distinct African languages.

To measure ethnic politicization, we employ group-level and individual-level indicators. At the group-level, we use the Ethnic Power Relations (EPR) (Vogt et al. 2015) and the Politically Relevant Ethnic Groups (PREG) datasets (Posner 2004a) to measure which ethnic groups or coalitions have been active in competition for state power in the post-independence period. To test ethnic salience at the individual level, we use data from Afrobarometer Rounds 3–6 that include a question on whether respondents self-identify more in ethnic or national terms. To

⁴See Caselli and Coleman (2013) for a formalization of the link between social closure and ethnicity.

analyze the hypothesized heterogeneous legacies of cash crops and print technologies on social closure and boundary-making, we employ a behavioural measure of ethnic assimilation: inter- and intra-ethnic marriages from a large sample of couples surveyed by USAID’s Demographic and Health Surveys. We also test whether the exclusionary ethnic boundaries induced by cash crop agriculture increased the risk of localized ethnic conflict compared to the inclusionary boundaries produced by print technologies.

We use linguistic groups identified in the Ethnologue database as our primary unit of analysis to minimize concerns about endogenous sample selection (Laitin 2000b, p. 142). This enables us to merge our cash crop, publishing, and outcome data, along with a host of control variables, to the Ethnologue groups through spatial overlays or ethnic name matching.⁵ In the survey-based analyses, we use two types of specifications. The first—*geographic-level* models—are based on the geographic location of individuals and the Ethnologue polygons in which they reside. These models compare people located in different places with and without historical cash crop production and/or missionary publishing. The second—*ethnic-level* models—are based on survey respondents’ affiliation to a given ethnic group rather than place of residence. Thus, they compare individuals residing in the same location but from ethnic groups with differential exposure to historical cash crop production and/or missionary publishing; this enables separating culturally transmitted attitudes and behaviors from locational effects.

We employ three main methods to mitigate endogeneity concerns. First, we employ location fixed-effects in our ethnic-level specification to address mission selection into areas with favorable locational fundamentals (Jedwab, Selhausen, and Moradi 2018). Second, we use our African-language publishing data to analyze the effects of print technologies at the intensive margin (i.e., estimating the effects of the magnitude of publication records among groups with at least one publication). Third, we instrument actual crop production with agro-climatic suitability to address the potentially endogenous uptake of commercial agriculture. We also conduct additional robustness checks to rule out alternative explanatory mechanisms.

Consistent with our pre-registered hypotheses, we find that groups historically exposed to cash crops or print technologies are significantly more likely to be politically-relevant after independence. Using the Ethnic Power Relations dataset, we find groups that cultivated one of five major cash crops through the end of colonialism or with a historical publication in their language are, respectively, 54 (125) and 43 (86) percent more likely than the average group to be politically relevant according to EPR (PREG). These results are robust to using cash crop

⁵For combining Ethnologue groups with information from EPR, PREG, DHS, and Afrobarometer, we use the publicly available ethnic links coded by Müller-Crepon, Pengl, and Bormann (2020).

suitability as an instrument and when focusing only on the sub-sample of groups exposed to Christian missions.

At the individual level, we find citizens residing in areas of historical cash crop production or living in Ethnologue polygons with a history of publishing are significantly more likely to self-identify with their ethnic group rather than nationality. Moreover, ethnic salience follows our expectation of cash crops producing location-specific effects among “stayers” and publishing producing broader cultural effects, including among “movers”, i.e. respondents living outside of their ancestral ethnic homeland. We do not find evidence, however, that groups treated with cash crops or print technologies have more homogeneous political preferences today.

We find strikingly different effects of cash crops and publishing on the porosity of ethnic boundaries, as measured by observed inter-ethnic marriage rates. Consistent with our expectation that cash crops engendered social closure and less openness to ethnic outsiders, we find inter-ethnic marriage to be significantly lower even with linguistically closely related groups. In contrast, and consistent with the hypothesis that print technologies led to salient but more porous ethnic boundaries, we find null effects on inter-marriage with linguistically close ethnic outsiders but negative effects on marriage across large linguistic distances. However, in contrast to our expectations, both exposure to cash crops and print technologies are positively associated with contemporary ethnic-based conflict—suggesting that, even as print technologies opened the door to assimilation of culturally proximate outsiders, its politicizing effects ensured these groups have not escaped cycles of ethnic conflict.

Our findings have important implications for a number of different research programs and bodies of scholarship in the social sciences. Despite the emergence of a strong consensus on the constructivist nature of ethnicity (Laitin and Posner 2001; Chandra 2012), the endogenous sources of ethnogenesis remain understudied. Our paper illuminates how Africa’s ethno-political landscape and boundary configurations were shaped by the impact of the cash crop and print revolutions from the 19th century onward. In doing so, we go beyond the role of border demarcation (Michalopoulos and Papaioannou 2016) and indirect rule (Posner 2005; Ali et al. 2019; McNamee 2019) in understanding social and economic processes of cultural change during the colonial era.

In pursuing this line of inquiry, we apply classic theories of group formation to Africa—Weber’s (1978) notion of social closure and Anderson’s (2006, pp. 46-47, 7) framework on the ethno-national impact of print technologies. To date, there have been few systematic tests of Anderson’s “imagined communities” hypothesis.⁶ Across Africa, we find strong support for the

⁶However, see Sasaki (2017), who focuses on the impact of the printing press in Europe on language standard-

socializing effects of the printing press and publishing in cultivating culturally persistent ethno-national identities that are open to incorporating linguistically-related ethnic outsiders—but which are more exclusionary of linguistically distant groups. However, as is well-known, unlike in Europe, this cultural change did not lay the foundation for nation-states. Instead the emergent ethno-national communities would be locked into externally-demarcated colonies arising from European imperial conquest.

Finally, our paper employs a pre-registered design to address growing concerns about publication bias and data mining for significant results in historical persistence studies. Beyond the benefits of guarding against cherry-picking positive findings, pre-registration encourages careful *ex ante* theorizing and hypotheses development. Pre-registration does not preclude *ex post* modifications of the pre-specified analyses, but it does necessitate transparency about any changes made. In this vein, we have included a table in the Appendix [A1](#) describing what we reported in the pre-registration and how we implemented the empirical analysis for the paper.

Conceptual Framework: The Determinants of Africa’s Ethnic Landscape

In this section we more fully advance our theoretical argument on the impact of the cash crop and print revolutions on shaping Africa’s modern ethnic landscape. Before addressing each in turn, we first situate our argument within the broader ethnicity scholarship.

Ethnic boundary-making and politicization

We conceive of a country’s ethnic landscape as shaped by two key processes: boundary-making and politicization. The former accounts for the process of group formation and change (ethnogenesis) — that is, how the construction and enforcement of social boundaries mediate a group’s shape and structure (Barth 1969; Wimmer 2013). Boundary-making is a function of demarcating criteria for group membership and policing those criteria by in-group members. Membership credentials or markers can be more or less accessible to outsiders (e.g., ascriptive characteristics compared to language) (Fearon 1999; Chandra 2006). The credentials salient at a given time and their enforcement have important implications for group size and change (Wimmer 2013). In contrast, ethnic politicization encapsulates the process by which some groups, but not others, serve as the bases for access to political and economic benefits (Bates 1983; Fearon 2006). It

ization.

entails large number of individuals consciously or subconsciously coordinating their political and economic behavior based on shared ethnicity.

What factors drive ethnic boundary-making and politicization? Existing literature on the origins of ethnic diversity emphasize four broad streams. One prominent sociobiological framework draws inspiration from Berghe (1981); it attributes Africa's higher levels of ethnic diversity to the founder effect—the loss of genetic variation as human species migrated from the cradle of humankind (Ahlerup and Olsson 2012; Ashraf and Galor 2013). A second stream emphasizes the role that geographic and environmental variation has played in producing cultural differentiation and enduring social divisions (Nettle 1998; Michalopoulos 2012). A third points to the importance of political institutions, especially a history of political centralization (Bockstette, Chanda, and Putterman 2002; Gennaioli and Rainer 2007; Ahlerup and Olsson 2012; Michalopoulos and Papaioannou 2013; Wimmer 2016), or their breakdown due to, for example, the slave trades (Nunn 2008). A fourth centers on the impact of colonialism. European colonization is argued to have reconfigured the ethno-political landscape via indirect rule (Mamdani 1996; Berman 1998; Posner 2005; Ali et al. 2019; McNamee 2019), the imposition of sovereignty and territorial partition (Englebert, Tarango, and Carter 2002; Posner 2004b; Michalopoulos and Papaioannou 2016), and differential access to modernizing benefits (Bates 1983; Horowitz 1985). We advance a fifth stream—drawing on a canonical literature on the processes of social closure (Weber 1978; Parkin 1979) and imagined communities (Anderson 2006)—that focuses on the economic and technological bases of ethnogenesis and politicization arising from the commercialization of agriculture and the spread of writing and printing.

The Cash Crop Revolution

In the 19th and 20th century, African economies underwent an important structural transformation away from the slave trades that dominated exchange for the previous four hundred years to commercial export agriculture (Hogendorn 1969; Hopkins 1973; Frankema, Williamson, and Woltjer 2018). The cash crop revolution led to an important spatial shift in economic production to areas suitable for oil palm, groundnuts, cocoa, coffee, and cotton as the loci of the region's new economic geography (Hopkins 1973). Fuelled by European-financed transportation infrastructure before and during colonialism, these cash crop zones were vertically integrated with export markets but with weak horizontal linkages with the rest of the colony (Rodney 1972; Hirschman 1977).

We posit that the spatial disparities arising from the cash crop revolution had important path dependent effects on ethnic politicization. The takeoff of export agriculture endowed

some groups—those who would be the primary producers of cash crops or the owners of the land on which they were produced—with a common economic niche and much greater wealth potential than others. Beyond its impact on the group’s mobilizational capacity, these groups are more likely to have had much more intense interactions with the colonial state as it penetrated their homelands to build infrastructure and develop markets. In some cases, such as in cotton enclaves in Central African Republic or Chad, state penetration was especially oppressive, perhaps hardening societal identities vis-à-vis the state (Lemarchand 1980).

A second important channel was via the effects of the commercialization of agriculture on land tenure regimes.⁷ Many of the most suitable areas experienced an increase in demand for land as waves of farmers, including enterprising migrant farmers (Hill 1963), adopted cash crops. Further increasing population growth and land pressure were the influx of migrant laborers, especially after the advent of colonialism and forced labor policies (Cordell, Gregory, and Piché 1996).

The commercialization of agriculture combined with migration-led population growth induced important changes in the social bases of land tenure regimes. In pre-colonial Africa, land rights were contingent on group membership or allegiance to traditional authorities (Berry 1993). These practices did not change per se with the advent of cash crop agriculture and colonialism. What did, however, were outsiders’ eligibility for group membership as ethnic boundaries became more tightly regulated (Boni 2006; Lentz 2013). Thus, following from Weber’s (1978) idea of social closure (Parkin 1979), in which social identity is employed as a means to restrict access to economic rents, in the face of rising land values and an in-flux of migrants, ethnic boundaries were more firmly policed to exclude outsiders from land ownership.⁸ In line with the idea that ethnic differences are constructed, at least partially, as “a boundary-enforcement device” (Caselli and Coleman 2013, 162) (see also, Fearon 1999) to control private goods, contestation over land not only made ethnicity more salient, it likely led sons-of-the-soil to emphasize less accessible criteria of group membership (such as, ascriptive characteristics and ancestral-ties to the land). In a fascinating ethnography of the impact of the spread of cocoa and migrant farmers to the Sefwi homeland (located in present-day western Ghana) from the early 20th century onward, Boni (2006) documents this precise dynamic unfolding—resulting in the “ancestralization of land rights” and more stringent enforcement to prevent migrants from permanently

⁷For important previous work, see Berry (1993) and Boone (2014, 2017).

⁸This process of ethnic boundary hardening was driven from below—as chiefs found themselves under growing pressure from their constituents not to give away too much land to outsiders (Boni 2006)—but also supported from above—as colonial government’s promoted neocustomary land tenure regimes (Boone 2017; Mamdani 1996, 104-105).

owning land. One potential legacy of the emergence of more exclusionary ethnic boundaries is that cash crop enclaves would become flashpoints for communal violence as impermeable social boundaries increased the rents that would come from violent appropriation (Caselli and Coleman 2013).

Christian Missions and the Spread of Writing and Print Technologies

As the abolition of the slave trade ushered in cash crop agriculture in Africa, it also gave momentum to the spread of Christian missions across the continent. In their endeavor to spread the Gospel, missionaries set in motion a communication revolution.

Many missionaries, especially Protestants, valued translating the Bible and education materials into vernacular languages as a vehicle for conversion (Laitin 2007; Woodberry 2012). As most African languages were oral languages or written in non-Roman script, missionaries first invested in developing Romanized writing systems (Posner 2003). Accordingly, the transformation introduced by missionaries in 19th and 20th century Africa was much more radical than the European print revolution studied by Anderson (2006), as it involved printing, but also entailed language categorization, standardization and transcription—mainly (though not exclusively) by non-native speakers (Ranger 1989; Chimhundu 1992). To propagate language knowledge and consumption of the written texts, printing presses were imported to mass produce texts that were then used in churches and schools (Posner 2003; Cagé and Rueda 2016).

The Yoruba represent a paradigmatic case of the impact of missionary language investments and publishing on the reconstruction of ethnic identity⁹—with countervailing social effects of those wrought by the slave trades.

With the collapse of the Oyo empire at the end of the 18th century, civil wars and slave raiding divided the Yoruba into rivalrous subgroups (Adediran 1984). From the 1840s onward, however, a group of missionaries from the Anglican Church Missionary Society (CMS), including freed slaves, such as Samuel Crowther, contributed to the rebuilding of the Yoruba ethnic nation. Intent on spreading Christianity to Yoruba-speakers, the CMS missionaries made considerable investments in Yoruba orthography, translation, and publishing (Falola 1999). In propagating a standardized language and embracing and promoting the ethnonym ‘Yoruba’, the Christian missionaries boosted Yoruba ethnic *consciousness* (Peel 2003). Moreover, as missionaries interpreted Yoruba history, culture and tradition through a Christian lens (most famously

⁹For other case studies, see Ranger (1989), Chimhundu (1992), and Strommer (2015).

Samuel Johnson in *The History of the Yorubas*), ethnogenesis and religious change reinforced each other.

Consistent with the Yoruba case, the standardization and printing of African languages is expected to have strengthened groups' ethno-nationalism and their mobilizational capabilities. We expect this to work via an Andersonian process of socialization as members collectively identify as part of an "imagined community," potentially reducing collective action costs that hinder political mobilization (Miguel 2004). Moreover, writing and printing equipped these groups with the mobilizational resources they needed to effectively compete in the political arena. The Yoruba again are an apposite case. With their defined identity, now-homogeneous language, written references to a common history, and newspapers and tracts, they emerged as a major player in contemporary Nigerian politics (Adesoji 2017).

In sum, publishing likely contributed to the emergence of more expansionary and inclusive identities than cash crop agriculture as missionaries encouraged language propagation and provided opportunities for outsiders to learn the language—via dissemination of language materials, church-related activities and schooling. This contributed to the construction of more porous ethnic boundaries and assimilationist cultural practices—at least among those who adopted the group's language.

Hypotheses

Following from our theoretical framework, we pre-registered a number of hypotheses on the effects of cash crop agriculture and print technologies on ethnic politicization and boundary-making.

In terms of politicization, we hypothesized that groups exposed cash crops or print technologies are more likely to be politically relevant in the post-independence period. We also expected this to lead to more salient ethnic identities among individuals from these groups and more homogeneous political preferences among group members.

Despite these similar effects on ethnic politicization, we expected differential effects on boundary-making. We hypothesized that the commercialization of agriculture led to the construction of less porous ethnic boundaries compared to the more open ones arising from publishing. Consequently, we predicted lower rates of inter-ethnic marriage and inter-group trust and higher rates of ethnic conflict for the cash crop than for the publication treatment.

Finally, we pre-specified that cash crops and print technologies would affect the salience of ethnicity through different channels. As cash crops' effects on ethnicity operated through its instrumental use in contests for local economic benefits, we expected salience to be tied to

locations where historical cash crop cultivation occurred. In contrast, if publishing led to the cultural transmission of ethnic identities through language standardization and propagation through printing, we expected salience to be transmitted across generations even as individual members (or their ancestors) moved outside their ethnic homeland.

Data

In this section we describe the various sets of data we assemble to test our hypotheses. We explain the use of Ethnologue to derive units of analysis, describe our historical data on cash crops and African language publishing, and discuss our proxies for ethnic politicization, salience, and boundary-making.

Historical and geographical data

Identifying potentially relevant groups. For a candidate list of nominal ethnic categories, we use Ethnologue, a reference source on living languages. Ethnologue aims to capture the *complete universe* of languages regardless of their social or political relevance, or demographic size (Simons and Fennig 2017). Having been compiled from the 1950s onward, Ethnologue may nevertheless miss a few pre-colonial small or extinct ethno-language groups. However, selection issues seem minimal in comparison to datasets like AMAR, EPR, or Murdock (1959, 1967).¹⁰ Identifying potentially salient ethnic categories from Ethnologue restricts our focus to ethno-linguistic rather than racial, religious, or regional markers. The analytical consequences of this restriction are minimal since in our Sub-Saharan African sample practically all ethnic categories in EPR, PREG, Afrobarometer, and DHS are equivalent to, or combinations of, language families, languages, or dialects. Another advantage of Ethnologue is that its companion dataset, the World Language Mapping System (WLMS) provides maps demarcating linguistic homelands, which we leverage to spatially aggregate our cash crop data, survey-based outcome measures, and geographic control variables as described in detail below.

Cash crops. To measure cash crop production, we use a geospatial dataset on the primary commodity revolution in Africa, drawing from a map produced by Hance, Kotschar, and Peterec (1961). The map depicts the source location of more than 95 percent of exports in 1957 across

¹⁰AMAR and EPR rely on some indication of social or political relevance as a basis for inclusion. Murdock (1959, 1967) has a much smaller number of groups than Ethnologue. See Laitin (2000b, p. 142) on the advantages of using “language as a proxy for ethnicity.”

38 states in sub-Saharan Africa.¹¹ Each primary commodity production point represents a value of \$289,270, in 1957 USD. The dataset covers 9 groups of cash crops¹², 20 minerals and metals, and forest, animal and manufactured products. Figure 1 maps the distribution of the commodity production points. The analysis below focuses on the five main cash crops: cocoa, coffee, cotton, palm, and groundnut representing 80% of total cash crop production and no less than half of all exports in 1957 across the countries in our sample.

Print technologies and publishing data. As noted above, Africa’s printing revolution was fueled by missionary publishing in vernacular languages for conversion. To capture exposure to this technological change, we draw from two library databases to construct a record of historical publishing.¹³ This represents the first geocoded database of publishing in African languages throughout the colonial period and after independence.

The first source is a 1923 compilation of 2480 publications across 168 languages (Rowling and Wilson 1923). It was intended to serve as a reference book for publications by Christian missionaries in Africa, including not just religious texts, but also dictionaries, grammar books, educational materials, and newspapers. An estimation of the number of language speakers is given for each language.

The maps in figure A1 in the Online Appendix show the number of publications per 10,000 language speakers and the share of published materials that are educational or linguistic (grammars and dictionaries). These maps also illustrate the overlap between missionary regions of influence and cash crop regions. Figure A2 in the Online Appendix shows the distribution of publications across categories.

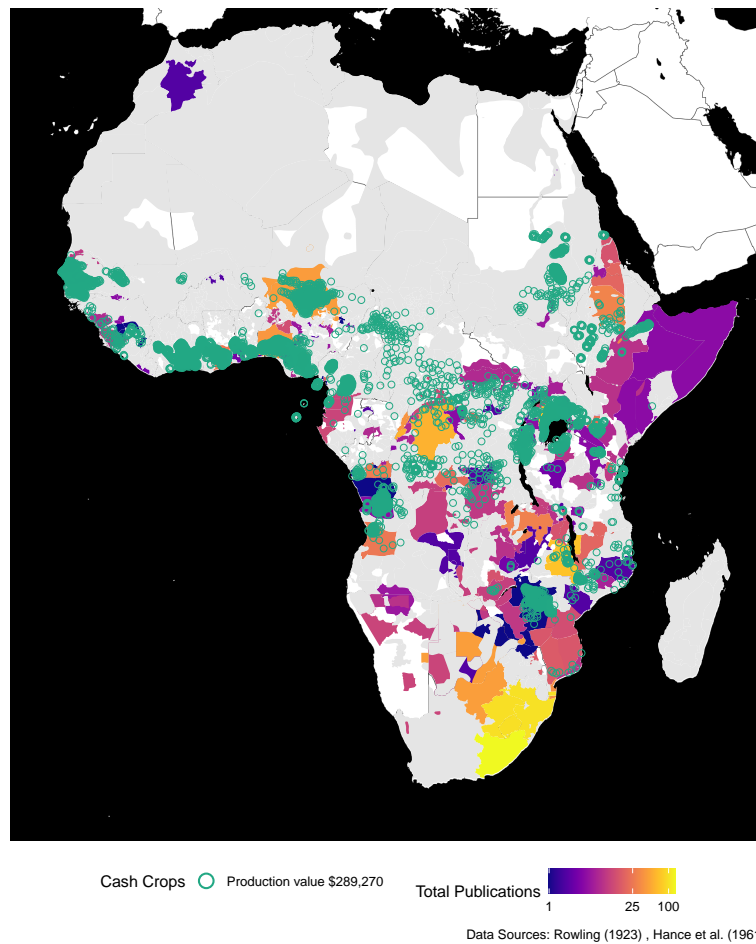
The second source, Mann and Sanders (1994, p. vii), “represents the combined collections of African language texts at SOAS (the School of Oriental and African Studies), (...) the African Department of SOAS, the International Institute for African Languages and Cultures (...) and the International Committee on Christian Literature for Africa.” This source complements Rowling and Wilson (1923) given its greater temporal coverage and the inclusion of texts produced outside of the Christian sphere of influence. However, Mann and Sanders (1994) exclude grammars and dictionaries, which may have been particularly important for constructing salient ethno-linguistic communities. It is therefore much less comprehensive on early printed materials, as it counts only 1270 pre-1925 publications (50 % less than Rowling and Wilson (1923)).

¹¹It excludes data on the Union of South Africa (including present-day Namibia), Madagascar and other island colonies.

¹²Cocoa, coffee, cotton, groundnuts, oil palm, stimulants, other food crops, other industrial crops, other oils

¹³This approach was inspired by Chaney (2016)’s work on the Middle East.

Figure 1: Publications and Cash Crop locations



Notes: Language homelands are mapped according to Ethnologue. Greyed regions are Ethnologue polygons for which there is no record of publications. The colors indicated the number of publications listed in Rowling and Wilson (1923). Each green circle locates 289,270 USD (1957) cash-crop export value.

We thus use Rowling and Wilson (1923) as the main source in our analysis; results using Mann and Sanders (1994) are presented in the Online Appendix.

The map in figure 1 shows the total number of publications for Ethnologue ethnic homelands cited in Rowling and Wilson (1923) and the location of cash crop production. The same map using the alternative Mann and Sanders (1994) data is presented in Appendix Figure A3.

Contemporary data on ethnic identities and political relevance

We use several data sources to measure the main outcomes of our study: ethnic politicization and boundary-making.

Group-level politicization measures. To measure which Ethnologue groups serve as bases for contemporary political mobilization, we match Ethnologue to two sources on ethno-political relevance: the Ethnic Power Relations (EPR) (Vogt et al. 2015) and the Politically Relevant Ethnic Groups (PREG) (Posner 2004a) datasets. For each, we code two types of binary variables indicating (a) “Exclusive links”: whether an Ethnologue group is listed under the same or a clearly synonymous name in EPR or PREG (e.g., Yoruba and Yoruba), and (b) “Links”: whether the group is a clearly identifiable part of a broader ethnic coalition coded as relevant by EPR or PREG (e.g., Gikuyu as part of the Kikuyu-Meru-Embu coalition).¹⁴

Individual-level politicization measures. Even among politically-relevant groups, there likely exists significant variation in the salience of one’s ethnicity vis-à-vis other identities, and how cohesively members behave in the political sphere. To analyze this, we use survey data from rounds 3–6 of Afrobarometer.¹⁵ Among other questions on citizen attitudes on democracy and governance, surveyed individuals are asked whether they identify more in ethnic or in national terms—which is commonly used as a measure of ethnic salience (Eifert, Miguel, and Posner 2010; Robinson 2014; Ali et al. 2019). For homogeneity of political preferences, we use Afrobarometer responses on vote intention to calculate a Herfindal index.

Boundary-making. As discussed above, a key dimension of boundary-making is a group’s accessibility to outsiders. Given the importance of marriage in social relations and group maintenance, many scholars view “endogamy [as] the ultimate measure of the salience of boundaries for intergroup relations” (Hechter 1978, 304). The underlying assumption is that groups with more exclusionary boundaries are less likely to marry outside their group—and to develop norms against such practices.

To calculate inter-ethnic marriage rates, we use the household component from USAID’s Demographic and Health Survey (DHS) that includes data on the ethnicity of individuals and their spouses.¹⁶ These measures are described in more detail below. More exclusionary group

¹⁴We also use AMAR (All Minorities At Risk) to measure groups’ social relevance capturing group consciousness and shared norms and cultural features short of national-level political mobilization (Birnir et al. 2015, 112).

¹⁵The Afrobarometer 3–6 sample includes the following countries: Algeria, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Côte d’Ivoire, Ethiopia, Gabon, Ghana, Guinea, Kenya, Lesotho, Liberia, Madagascar, Mali, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Sao Tome and Principe, Senegal, Sierra Leone, South Africa, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe.

¹⁶The DHS are standardized nationally representative surveys in developing countries, including the following sets of African countries: Angola, Benin, Burkina Faso, Burundi, Cameroon, Central African Republic, Côte d’Ivoire, DR of Congo, Ethiopia, Gabon, Ghana, Gambia, Guinea, Kenya, Lesotho, Liberia, Madagascar, Mali, Malawi, Mozambique, Namibia, Niger, Nigeria, Sao Tome and Principe, Senegal, Sierra Leone, South Africa, Swazi-

boundaries might also manifest in differential levels of trust. To capture this, we use Round 3 of the Afrobarometer surveys to measure what Robinson (2017) calls “ethnocentric trust”—in which respondents report trusting co-ethnics more than non-co-ethnics or people in general. Questions on social trust are limited to round 3 of the Afrobarometer. Given this and their vulnerability to social desirability bias, we focus our boundary-making analysis on inter-ethnic marriage.

Cash crops, print technologies, and politicization

Following our pre-analysis plan (see Appendix A1), we first report the results of the effects of cash crops and publishing on ethnic politicization at the group and individual levels.

Group-level Specification and Results

To test group-level effects, we estimate regression equation 1 using OLS.

$$\text{Pol}_{ec} = \beta_0 + \beta_1 \text{Cash crops}_{ec} + \beta_2 \text{Publications}_e + \mathbf{X}'_{ec} \boldsymbol{\gamma} + \lambda_c + \varepsilon_e \quad (1)$$

Pol_e measures the political relevance of Ethnologue group e in country c , using EPR or PREG. Cash crops_{ec} is a binary measure of historical cash crop cultivation in the Ethnologue polygon. Publications_{ec} indicates at least one historical record of publishing is listed in Rowling and Wilson (1923). λ_c represents country fixed-effects. \mathbf{X}_{ec} is a set of standard geographic and historical controls (aggregated to WLMS polygons), including agricultural suitability, tsetse fly and malaria ecology, elevation, ruggedness, average yearly precipitation, average yearly temperature, agricultural suitability, distances (in logs) to the coast, to navigable rivers, to cities in 1900, to the country capital, to a historical missionary settlement, and to missionary printing presses, as well as absolute longitude and latitude.

Figure 2 reports the estimates of regression 1 when the outcome is a binary variable equal to one if the ethnic group is listed as politically relevant in EPR or PREG. Focusing on columns 1 and 3 of Figure 2, the results indicate that, conditional on controls, a group with historical cash crop production is roughly 16-17 percentage points more likely to be listed as politically relevant in EPR and PREG (a 54% and 125% increase from the sample mean of the dependent variable, respectively). Similarly, a group speaking a language with historical publishing is 11-12

land, Tanzania, Tchad, Togo, Uganda, Zambia, Zimbabwe.

percentage points more likely to be listed as politically relevant in EPR and PREG (a 53 and 85% increase from the respective outcome mean). We also find evidence that these historical forces increased groups' social relevance as measured by AMAR, an important antecedent to politicization (see Figure A6).

Potential endogeneity necessitates caution in causally interpreting the correlations reported in Figure 2. One paramount concern is the results are driven by geographic or historic confounds of ethnic groups that take up cash crops and print publishing.¹⁷

We employ several strategies to address this issue. First, we instrument “Cash crops_{*e*}” with indicators of suitability for cash crop agriculture and estimate the effects using a spatial-2SLS (S2SLS) strategy, following Betz, Cook, and Hollenbach (2019). The instrument is the average agro-climatic suitability from the FAO GAEZ database across the five most important African cash crops (cocoa, coffee, cotton, groundnuts, and oil palm) in the homeland of ethnic group *e*. These suitability scores combine various soil and climatic characteristics to predict the ecological potential to grow specific crops in rainfed agricultural systems. To serve as a valid instrument, suitability may only affect outcome variables through its impact on actual cash crop production. We argue that this exclusion restriction likely holds conditional on the rich set of geographic and historical controls in our models, especially general agricultural suitability, temperature, and precipitation which are included to isolate cash crop-specific effects from overall agricultural productivity and its social and political consequences.

The suitability instrument strongly predicts late colonial cash crop production in first-stage regressions. The first-stage F-statistic is 17.9 in the EPR and 17.7 in the PREG models. To account for potentially similar spatial patterns in the instrument and outcomes that may threaten the exclusion restriction, the IV models further include a spatial lag of the respective political relevance outcomes instrumented with first and second-order spatial lags of the geographic and historical control variables (Betz, Cook, and Hollenbach 2019). All spatial lags are based on a binary contiguity matrix which defines ethnic group *e*'s neighbors as all other ethnic polygons within a 100 km centroid distance.¹⁸ Line 3 in Figure 2 shows that S2SLS results remain remarkably similar to baseline OLS although confidence intervals naturally become wider.

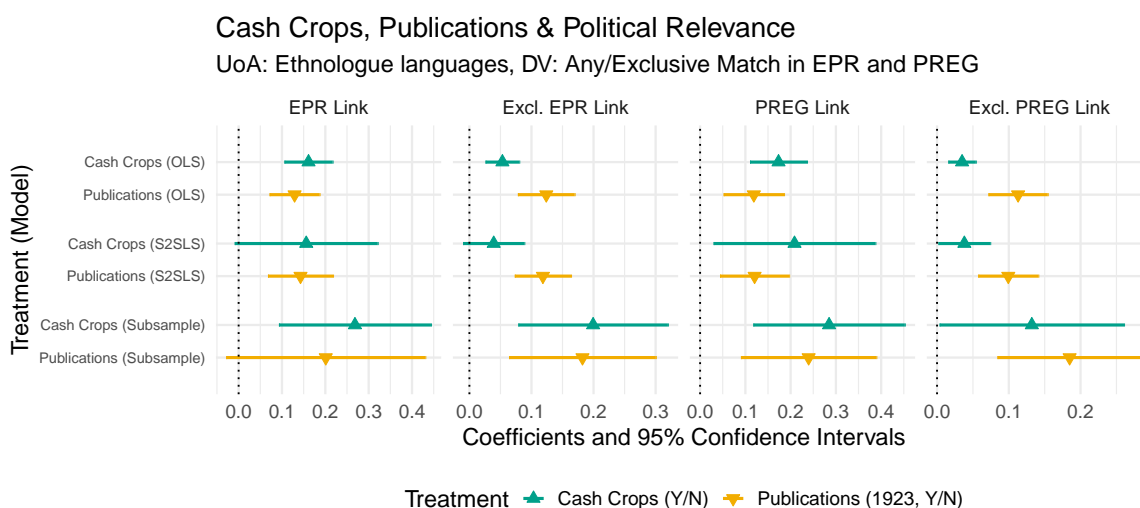
A second endogeneity concern is that European missions tended to establish outposts in geographically-favorable areas (Jedwab, Selhausen, and Moradi 2018). Subsetting the analysis to

¹⁷Appendix Figure A4 shows that groups listed as politically relevant exhibit statistically different standardized mean differences across covariates.

¹⁸The joint significance of spatially lagged baseline controls in the second first stage (predicting the spatially lagged dependent variable) is high. The respective F statistics are 32.8 (EPR Link), 43.0 (Excl. EPR Link), 33.0 (PREG Link), and 97.2 (Excl. PREG Link).

groups exposed to missions makes the analysis sample more comparable in terms of geographic fundamentals. The results, reported in lines 5-6 of Figure 2 are generally robust, despite the large reduction in observations (and thus larger standard errors). We also estimate 1 at the intensive margin—with cash crops calculated as the total value (in 1957 \$ US), and publishing measured as the number of historical publications per capita. We restrict the sample to ethnic groups with at least one historical publication in its language in Rowling and Wilson (1923) or Mann and Sanders (1994). Results are reported in Table A6 and A7 in the online Appendix. They are generally robust for exclusive links¹⁹. For non-exclusive links, the treatment effects retain the same magnitudes but lose precision..

Figure 2: Cash Crops, Print Technologies, and Political Relevance



Notes: These figures summarise the results of 12 regression models. 4 binary outcomes are used: “EPR Link” (resp. “PREG link”), which flags whether an Ethnologue group is non-exclusively matched to a group listed as politically relevant in EPR (resp. PREG), “Excl EPR Link” flags an exclusive (1-to-1) match between the same databases. Lines 1-2 report treatment effects using binaries indicating whether Ethnologue groups were exposed to cash crop production and/or print technologies. In lines 3-4, cash crops are instrumented with the mean agro-climatic suitability for the five most important export crops using the S2SLS method described in the text. In lines 5-6, the sample is restricted to Ethnologue polygons that experienced missionary activity.

Individual-level Specification and Results

To test individual-level effects, we use survey data of expressed ethnic salience and estimate the following regression equation:

¹⁹The results on the publication treatment lose precision when using EPR outcomes. The p-values increase to roughly 12%.

$$\text{Sal}_{ielcs} = \mu_0 + \mu_1 \text{Cash-Crop}_{kcs} + \mu_2 \text{Publications}_{kcs} + \mathbf{W}'_{ilcs} \boldsymbol{\gamma} + \eta_{k'} + \varepsilon_{ielcs}$$

$$k \in \{e, \ell\}, \quad k' \in \{\ell, cs\} \quad (2)$$

Sal_{iel} is a binary Afrobarometer-based survey measure of greater ethnic than national identification. The unit of analysis now is respondent i , who identifies with ethnic group e , residing in survey location ℓ in country c , and is interviewed in Afrobarometer survey round s . We assign Afrobarometer respondent i to ethnic group e based on the language they report speaking at home.²⁰ We further use geographic information on Afrobarometer’s survey locations ℓ to assign individual respondents to an Ethnologue polygon. The treatments are defined by cash crop production within a 15 km radius of a survey location and the publishing record of a given Ethnologue group. Fixed effects η are either at the country-round cs level for geographic specifications (when $k = \ell$) or at the survey location-level ℓ , for ethnic-level specifications (when $k = e$). Whenever present, geographic and historical controls are the same as those stated in the previous section. Individual-level controls include gender, age, education levels, and indicators of standards of living.

The results of the geographic-level analysis are reported in Table 1. They show that a one standard deviation increase in the value of cash crop production at location ℓ increases respondents’ ethnic identification by around 1.2% of a standard deviation (approx. 0.4 percentage points or 3% of the mean outcome) (columns 1 and 3). Similarly, a one-standard deviation increase in publications per thousand people increases ethnic identification by around 3.7% of a standard deviation (approx 1.3 percentage points or 10% of the mean outcome). The effects are driven by ethnic-stayers—individuals of ethnicity e who still reside in one of the Ethnologue polygons constitutive of that ethnic group. Column (4) indeed shows that, when focusing on ethnic-leavers (those who reside outside of their ethnic group’s homeland), the effect of print technologies drops closer to 0 and loses statistical significance, and the effects of cash crops becomes negative.

In the ethnic-level analysis, treatments are now defined via respondents’ self-reported ethnic affiliation e (cash crops and publications aggregated up to the ethnic group level using constitutive Ethnologue polygons) with fixed effects at the location level ℓ . The results are reported in Table 2. Among individuals within the same survey location, ethnic salience is significantly

²⁰As a robustness check, we measure ethnic identity based on an individual’s stated ethnicity (i.e. their response to the question “What is your tribe? You know, your cultural or ethnic group.”). We prefer the language-based measure because it has fewer missing observations.

Table 1: Geographical Persistence in Ethnic Identity

	Geographic- level - Ethnic vs National Id				
	(1)	(2)	(3)	(4)	(5)
Cash crops USD pkm2	0.013*** (0.004)		0.011*** (0.004)		-0.015** (0.007)
Cash crops USD pkm2 - S2SLS				-0.007 (0.006)	
Pubs pth pop (1923)		0.038*** (0.006)	0.037*** (0.006)	0.037*** (0.006)	0.003 (0.007)
Individual controls	yes	yes	yes	yes	yes
Historical and Geo controls	yes	yes	yes	yes	yes
FE	Country-Round	Country-Round	Country-Round	Country-Round	Country-Round
Ethnic Stayer/Leaver	Both	Both	Both	Both	Leaver
Mean dep. var.	0.1294	0.1294	0.1294	0.1294	0.1294
Observations	136,255	132,863	132,863	132,863	45,301
R ²	0.041	0.042	0.042	0.042	0.044

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports standardized OLS estimates (“beta” coefficients). Standard errors are reported in parenthesis and clustered at the location level. The dependent variable is a binary variable flagging whether respondent declares a stronger ethnic than national identity. In column (4), we instrument cash crop production with agricultural suitability to cash crop production using the S2SLS described in the text. In column (5) the sample is restricted to locations that are referenced in our bibliographies.

higher among those ethnic groups with a history of publishing. A one standard deviation increase in publications per (est.) thousand people increases respondents’ ethnic identification by around 0.9% of a standard deviation (approximately 0.3 percentage points, or 2.4% of the mean outcome) (columns 2 and 3). In contrast, historical cash crop production now has no significant effect.

Whereas cash crops increased ethnic salience only among stayers, publishing significantly elevates ethnic identities among movers (Column (4)). In contrast to the geographic analyses, this cultural mover effect is robust to intensive margin-only comparisons (Column (5)). This suggests a culturally transmitted effect of print technologies—the formation of an “imagined community”—which persists even as individuals (or their ancestors) migrate.

Overall, we find that ethnic groups with higher levels of historical cash crop production and publishing are more likely to be politically relevant in the post-independence period and that individuals from these groups report more politically salient ethnic identities. However, despite higher levels of political salience, this does not necessarily manifest in more homogeneous political preferences (at least as measured by self-reported vote intention; see Appendix Figure A7). The individual-level ethnic salience results suggest we are capturing two different channels of politicization—one tied to place and the other stemming from cultural transmission. That these correlate, respectively, with localized cash crop production and language transforma-

Table 2: Cultural Persistence in Ethnic Identity

	Ethnic-level - Ethnic vs National Id				In Biblio
	(1)	(2)	(3)	(4)	
Cash crops USD pkm2	-0.010 (0.007)		-0.012 (0.007)	-0.009 (0.010)	0.011 (0.015)
Pubs pth pop (1923)		0.008* (0.005)	0.009* (0.005)	0.015** (0.006)	0.014* (0.008)
Individual controls	Yes	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	No	No	No
Fixed Effect	Town	Town	Town	Town	Town
Ethnic Stayer/Leaver	Both	Both	Both	Leaver	Leaver
Mean dep. var.	0.1294	0.1294	0.1294	0.12	0.11
Observations	127,000	130,161	126,958	45,343	31,590
R ²	0.202	0.200	0.202	0.263	0.268

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports standardized OLS estimates (“beta” coefficients). Standard errors are reported in parenthesis and clustered at the location level. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity.

tion increase our confidence that these historical processes were at least part of the causal chain shaping ethnic identity across countries in Africa.

Cash crops, print technologies, and boundary-making

We now turn to analyzing ethnic boundary-making—principally operationalized through inter-ethnic marriage.²¹ We hypothesized that cash crops and publishing differently transformed ethnic boundaries: with the commercialization of agriculture leading to economic competition and social closure, resulting in a legacy of sharp ethnic boundaries; and publishing facilitating ethnic assimilation and more porous boundaries through the development of imagined communities.

To measure ethnic exogamy, we take advantage of the couple recodes of the DHS household surveys which capture self-reported ethnic identities of married couples. The empirical specifications are equivalent to the Afrobarometer-based geographic and cultural persistence models above, but now the unit of analysis is interviewed couple i residing in location ℓ in country c with spouses identifying with ethnic group(s) e_f and e_m . Knowing the appropriate match of practically all raw ethnic categories in DHS on the Ethnologue language tree allows us to analyze

²¹We also pre-registered an analysis on inter-ethnic trust using data from Afrobarometer from round 3. See Appendix Table A5. We find individuals from historical cash crop areas tend to have overall lower levels of trust, and, conditional on this, report being more trusting of their co-ethnics. Print technologies have no effect on self-reported inter-ethnic trust.

inter-ethnic marriages at different levels of ethno-linguistic differentiation.²² Ethnologue has 13 levels of language differentiation d in our Sub-Saharan African sample. Differentiation $d = 1$ distinguishes broad language families and as d increases, more closely related ethno-linguistic categories are separated. We therefore define 13 binary outcome variables Sal_{icel}^d indicating if the two spouses in respondent couple i self-report belonging to different ethnic groups at level of differentiation d .

Two examples from Nigeria illustrate the operationalization of our inter-ethnic marriage outcomes. A marriage between a female respondent identifying as Yoruba and a male Hausa respondent is coded as exogamous on all levels of the language tree. The Yoruba language belongs to the Niger-Congo language family, whereas Hausa is an Afro-Asiatic language. These language families are already separate on the first level and therefore Yoruba and Hausa do not share any nodes on the language tree. In contrast, a Yoruba-Igala couple is coded as endogamous on levels 1-6 of and as exogamous thereafter. The Yoruba and Igala languages share the first six nodes of the language tree (Niger-Congo/Atlantic-Congo/Volta-Congo/Benue-Congo/Defoid/Yoruboid) but then branch out in different directions (Edekiri/.../Yoruba vs. Igala/.../Igala).

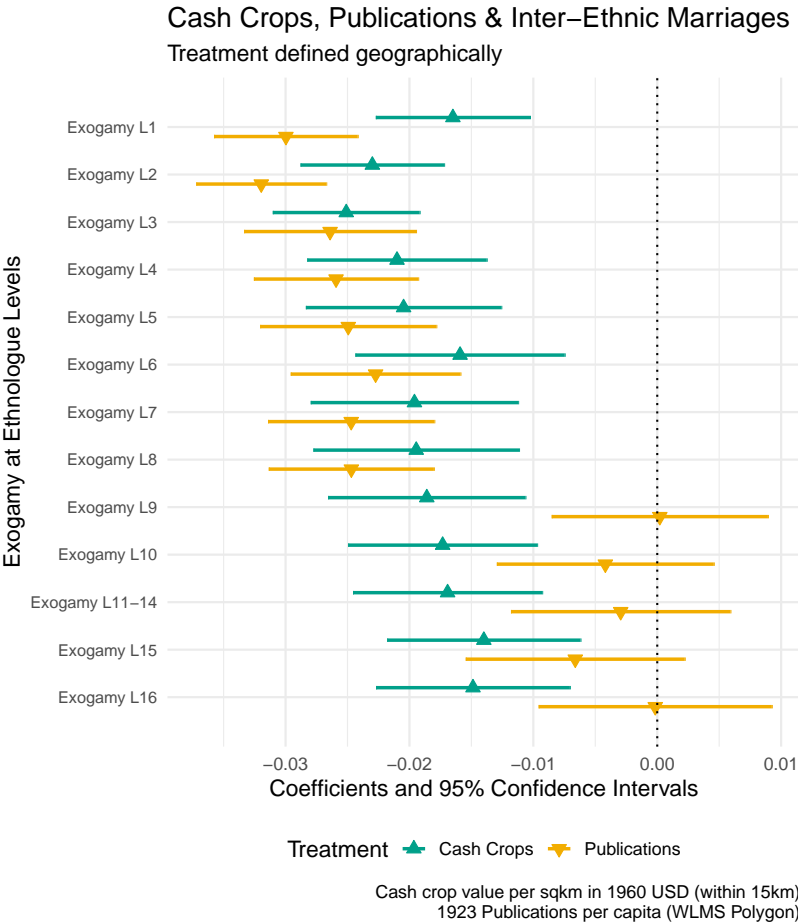
If cash crop agriculture sparked a process of more exclusionary identities, we would expect lower inter-ethnic marriage rates at even the furthest branches of the language tree. A Yoruba respondent from a cash crop region would be similarly less likely to be married to a Hausa as to an Igala speaker. If print technologies led to salient but porous ethnic boundaries, we would expect members of these groups (e.g., Yoruba) to be less likely to choose a spouse from a linguistically distant group (e.g. Hausa), but still open to inter-marry with linguistically related ethnic others (e.g. Igala).²³ We test these hypotheses for both the geographic and ethnic-level definitions of our treatment, as defined above.

Geographic Persistence. Figure 3 presents coefficient estimates from 13 models based on geographically assigned treatment variables. All 13 exogamy outcomes and both treatment variables are standardized to mean 0 and sd 1 to facilitate comparing coefficient sizes across Ethnologue levels and treatments. The cash crop coefficients in Figure 3 are consistently negative and significant across all linguistic levels of differentiation. Inter-ethnic marriages are between 0.013 and 0.02 standard deviations less likely in locations with one standard deviation higher levels of late colonial cash crop production. The coefficients on the publication variable are negative, significant, and somewhat larger in absolute size on levels 1–8 of the Ethnologue language tree.

²²see Cervellati, Chiovelli, and Esposito (2018) for a similar approach

²³The process of “absorption” of related ethno-linguistic groups as a result of language standardisation is also described for Europe by Anderson (2006) and further studied by Sasaki (2017).

Figure 3: Geographical Persistence - Cash Crops, Publications & Ethnic Marriages



Notes: The figure reports standardized OLS estimates (“beta coefficients”) from 13 regressions with country-round fixed effects. Standard errors are clustered at the survey location level. Each triangle represents the coefficient of geographic-level cash crop and print technology treatments, as described in the text. Bars represent 95% confidence intervals.

From level 9 onward, publication coefficients drop substantially and become statistically indistinguishable from zero. This pattern supports our theoretical conjecture that African language printing heightened the salience of ethnic identities but, compared to cash crop agriculture, led to more porous boundaries and more assimilation among linguistically close ethnic categories.²⁴

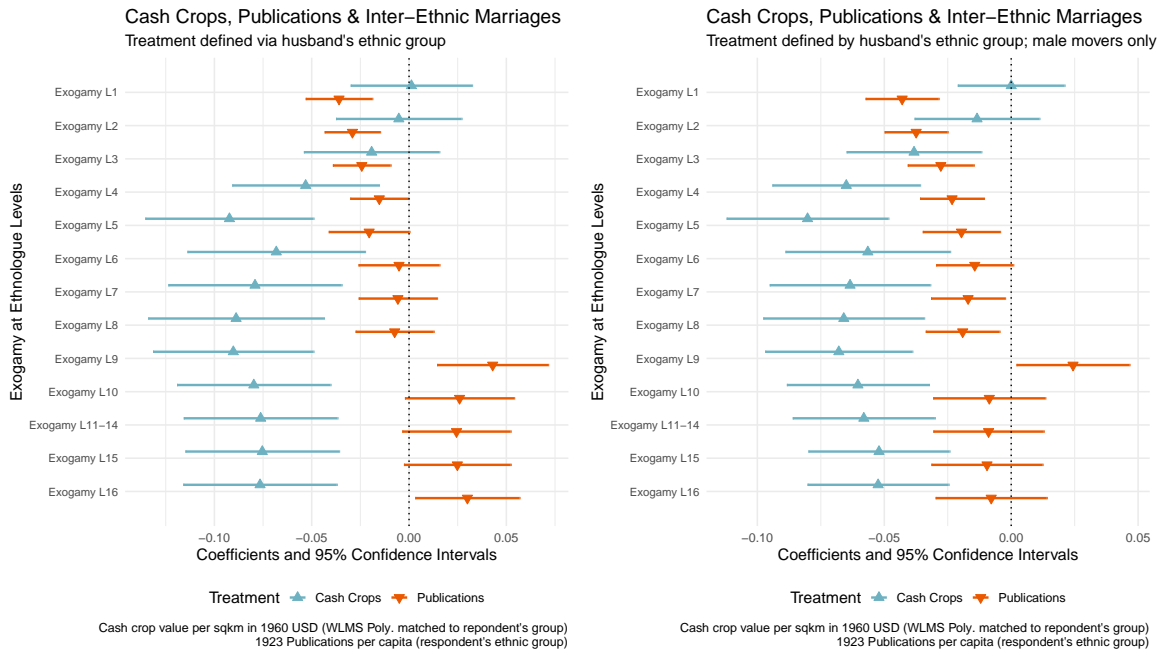
Appendix Figure A9 shows two alternative specifications. First, we define cash crops and all control variables at the ethnic polygon level rather than the more local DHS enumeration areas. Second, we use our alternative publishing data based on Mann and Sanders (1994). The results remain robust. Appendix Figure A10 replicates our baseline analyses on four subsamples of male and female ethnic movers and stayers. Results suggest that the effects of cash are driven by historically ‘native’ ethnic groups rather than respondents who (or whose ancestors) migrated to a given location exposed to cash crop production and/or missionary publishing. These findings are consistent with local sons or daughters of the soil being the main agents of ethnic boundary making.

Cultural Persistence. Figure 4 summarizes results from models that assign treatment variables by husbands’ ethnic identities and include location fixed effects. The left-hand panel in Figure 4 reports findings from analysis of the entire sample of couples for which both spouses’ ethnic identity was successfully matched to the Ethnologue language tree, whereas the right-hand panel restricts the sample to ethnic movers only and thus compares marital choices by husbands outside of their ancestral homeland. These within-location models yield substantively similar results as the geographic persistence analysis above. Effect sizes and the level difference between historical cash crop production and African language publishing appear, if anything, more pronounced. Appendix Figure A11 reports models that assign both treatments based on the wife’s instead of the husband’s self-reported ethnic group. The cash crop coefficients are comparable to the male movers analysis but the publishing effects are smaller and tend to lose statistical significance.

Addressing Endogeneity. We perform additional tests to address endogeneity concerns. First, we replicate the geographic persistence analysis for cash crops using agro-climatic suitability in reduced form and spatial instrumental variable specifications. The spatial IV analyses translate the logic of the group-level approach described above to the DHS setup with individ-

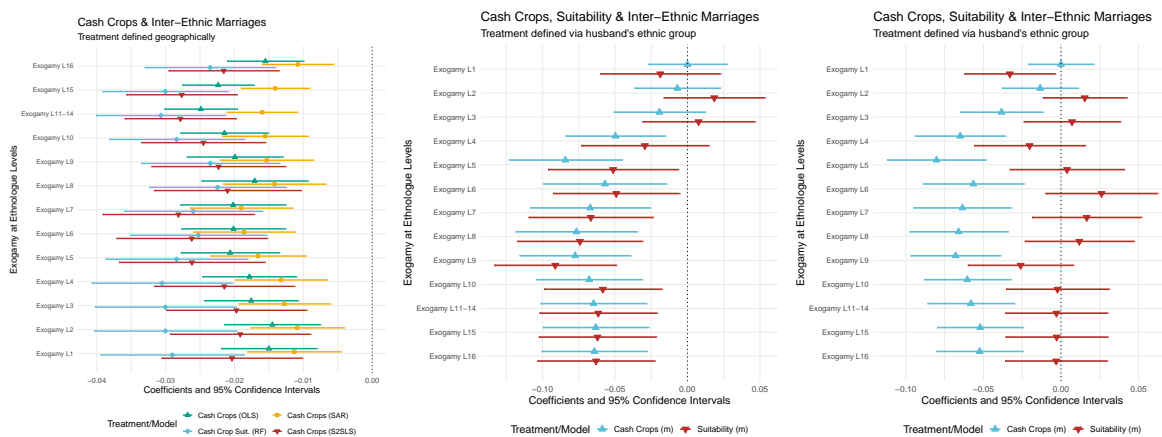
²⁴Appendix Figure A8 presents unstandardized specifications to facilitate interpretation of magnitudes. The effect sizes associated with a one standard deviation, zero-to-one, or 100% increase in the treatments of interest amount to 2.8-24.5% relative to the mean and are similar to logged distance to coast and a precolonial statehood dummy.

Figure 4: Cultural Persistence - Cash Crops, Publications & Ethnic Marriages



Notes: Each triangle represents the standardized OLS estimates (“beta coefficient”) of ethnic-level cash crop and print technology treatments, as described in the text.. The left panel is based on analyses of the whole sample while the right panel reports results from models run on the subsample of ethnic movers only. Bars represent 95% confidence intervals.

Figure 5: Endogeneity of Cash Crops?

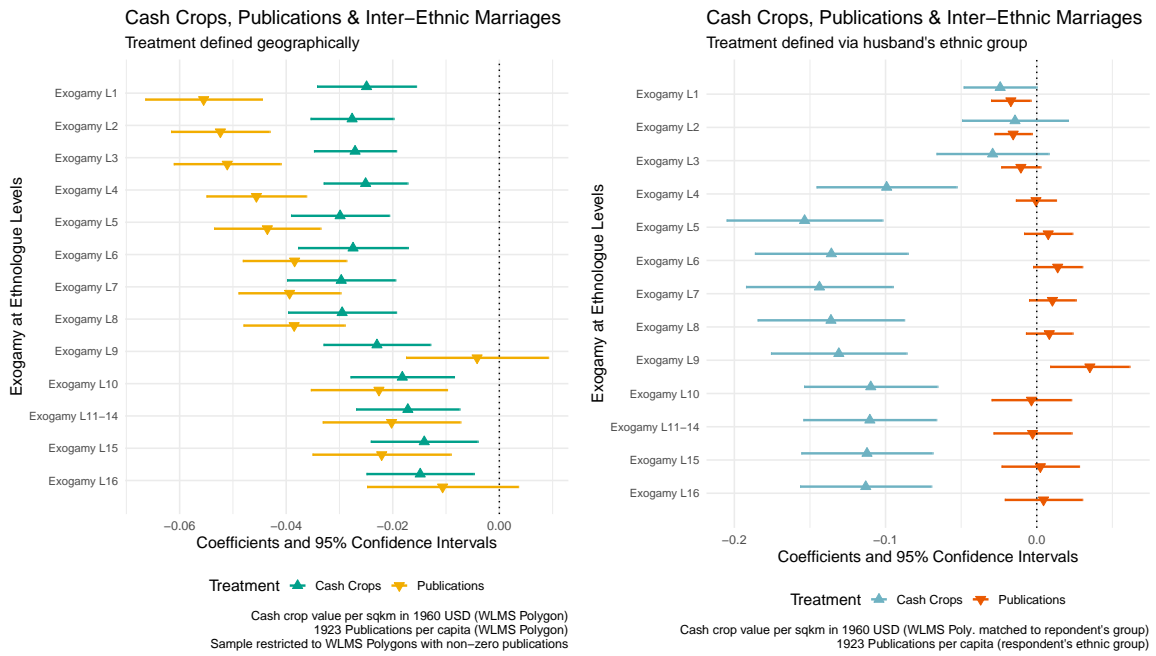


Notes: *Left:* The figure reports estimates from OLS, reduced form (cash crop suitability), spatial lag (estimated via 2SLS where the autoregressive spatial parameter is instrumented with first- and second-order spatial lags of the baseline controls), and spatial lag IV models (S2SLS method described in the text). The first-stage F-statistic of suitability is 47.54. The combined F statistic of the spatially lagged geographic controls ranges from 67.2 to 108.92. *Middle & Right:* The figures replicate the cultural persistence analysis from Figure 4 replacing the ethnically assigned cash crop variable with its suitability-based counterpart. As above, the sample in the right-hand panel is restricted to ethnic-mover husbands. All figures report standardized coefficients (“beta coefficients”).

ual couples nested in survey locations. We use mean cash crop suitability in survey location ℓ as an instrument and also include a spatial lag of the location mean of inter-ethnic marriages at Ethnologue level d instrumented as described above. The left-hand panel in Figure 5 presents coefficients from baseline OLS, spatial lag, reduced form, and spatial lag IV models. Reduced form and spatial IV estimates remain similar if somewhat larger than OLS, whereas coefficients in the spatial lag models get slightly smaller. The robustness of findings in these models suggest that the cash crop results are unlikely to be explained by unobserved confounding or spatially correlated outcomes (Kelly 2019).

The middle- and right-hand panel in Figure 5 summarize reduced form results alongside baseline OLS for the cultural persistence models shown in Figure 4. The within-location analyses of husbands with differential values of cash crop suitability in their ancestral homeland produces remarkably similar results when run on the complete sample but coefficients drop in size and lose significance when we restrict the sample to ethnic movers only.

Figure 6: Publications - Intensive Margins only



Notes: The figure presents results from intensive-margin only versions of the geographic persistence models from Figure 3 and the cultural persistence models from Figure 4 (right-hand panel). The left-hand side results are based on a sample restricted to WLMS polygons of languages with at least one publication in 1923; the right-hand panel uses a subsample of husband ethnic movers whose ethnic group had at least one publication in 1923. All figures report standardized coefficients (“beta coefficients”).

To address the potentially endogenous assignment of missionary language standardization and publishing activities to already mobilized ethno-linguistic groups, we run intensive margin only analyses. The left-hand panel in Figure 6 restricts the sample to DHS couples residing in WLMS polygons with at least one Christian text in our 1923 dataset. The publication coefficients are again negative and significant at higher levels of the language tree and get smaller in size but, this time, remain significant at more fine-grained levels of linguistic differentiation. The right-hand panel is based on a subsample of male ethnic movers whose self-reported ethnic group had at least one publication in 1923. While the basic pattern remains intact, coefficients become smaller and already lose statistical significance at and below level 4.

Discussion

Our empirical results from the previous sections suggest that: (i) historical cash crop production and uptake of print technologies and publishing increased groups' mobilizational capabilities and political relevance in the post-independence period; (ii.) these historical forces also have had persistent effects on individual ethnic salience but through different channels—cash crop effects appear tied to land and sites of historical cultivation, and publishing effects stem from cultural transmission among broader members of the ethno-linguistic group; and (iii.) we observe differential effects on patterns of inter-ethnic marriage, with exogamy less likely among those from cash crop groups at even the furthest branches of the language tree. Note that in contrast to the Afrobarometer models, we find cultural persistence (ethnic mover) effects of cash crops on ethnic marriages suggesting perhaps that political ethnicity is easier to change than deep-rooted cultural norms about appropriate marital choices. This seems consistent with recent findings that local ethnic minorities face incentives to vote for the local majority candidate rather than one of their own (Ichino and Nathan 2013).

Despite the differential social effects arising from cash crops and printing, both technologies appear to have left a legacy of local ethnic conflict, as reported in Appendix Figure A17. Employing a number of different measures of ethnic conflict and various methods to account for reporting bias due to better infrastructure and media penetration in these areas, geographic-level coefficients on historical cash crop and missionary publishing dummies are positive across all 30 specifications and statistically significant in all but three cases. The effects of cash crops align with Caselli and Coleman's (2013) framework on ethnicity as a technology of exclusion which increases incentives for conflict. What accounts for the positive and significant effects of publishing? One possible explanation is that, even as print technologies opened the door to the

assimilation of linguistically-related outsiders, this merely displaced conflict to higher-level ethnic cleavages—pointing to the potential violent consequences of strengthening imagined communities. Testing this mechanism versus others (such as, communal conflict as a downstream consequence of ethnic politicization or weakened state capacity due to stronger ethnic institutions) represents an important area for future research.

In the remainder of this section, we summarize findings from our pre-specified analyses to account for potential endogeneity before augmenting our pre-registration to address a series of potential alternative explanations that might account for the observed empirical patterns.

Addressing Endogeneity. Across most analyses, we address threats that the effects of historical cash crop production and vernacular language publishing are endogenous to underlying geographic factors or ethnic groups' pre-colonial characteristics. The effects of cash crops on group-level politicization (Figure 2) and inter-ethnic marriages (Figure 5) are robust to instrumenting cash crop production with indicators of suitability for cash crop agriculture and estimating the effects using a spatial-2SLS strategy.²⁵ To account for missionary and publishing selection effects, we show the results are robust to restricting the analysis to Ethnologue groups with a Christian mission (Figure 2) and publishing at the intensive margin (Tables 1 and 2, and Figure 6). To address potential geographic confounders of publishing, the results presented in Table 2 and Figure 4 include location-fixed effects. This increases our confidence that geographic confounders cannot explain away exogamy patterns or cultural persistence in ethnic identity.

Alternative Explanations It may be a worry that larger, more centralized groups were more intensively targeted by missionary printing and also more likely to engage in cash-crop agriculture. Table A12 and Figure A16 in the online Appendix show robustness to including pre-colonial ethnic centralisation from Murdock (1967) as a control in our regressions. We also show that the effects of cash crop agriculture and publishing on ethnic politicization and marriage patterns are not an artifact of British colonial rule (Ali et al. 2019). As reported in Appendix Table A13 and Figure A15, our results remain robust to inclusion of a British colony interaction term. Finally, other downstream consequences of cash crop agriculture and publishing, such as higher education or income, could also affect ethnic politicization, potentially in the opposite direction. Appendix Figure A12 and Table A10 show that the results are generally robust to

²⁵Afrobarometer results for cash crops are not robust to this instrumentation. One explanation is the lower spatial coverage of Afrobarometer which has less than half the number of unique survey locations than DHS in the subset of countries surveyed by both projects. In addition, Afrobarometer was geocoded ex post and location coordinates are probably less accurate than in DHS.

dropping these potential “bad controls”.²⁶

On boundary-making, one concern about the inter-ethnic marriage results is whether they merely reflect differences in local ethnic diversity. Appendix Figure A13 interacts both cash crops and publications with local-level ethnic fractionalization scores. If anything, our results are stronger in more ethnically diverse locations. This further strengthens our confidence that ethnic competition rather than local-level ethnic homogeneity explains lower exogamy in historical cash crop areas.

An alternative account of our results on exogamy is that members of ethnic groups exposed to print-technologies are also more likely to be Christian, and may be less open to partners with different from different religions. To rule this out, we re-run all baseline exogamy models with directed religious couple fixed effects that only exploit variation between couples with exactly the same religious denomination(s) between wife and husband. Results remain practically indistinguishable from our baseline analyses (See Appendix Figure A13).

Conclusion

Our analysis across 35 countries shows that Africa’s contemporary ethnic landscape was powerfully shaped by the persistent effects of the cash crop and printing revolutions that spread from the 19th century onward. In line with our pre-registered hypotheses, geographic variation in factor endowments for cash crop agriculture and the uneven diffusion of print technologies differentially increased groups’ mobilizational potential and their capabilities to compete for state power after independence. As anticipated, the survey-based analyses suggest that these two forces impacted ethnicity through different channels—with cash crop effects on individual identity salience tied to historic agricultural zones and publishing effects transmitted culturally among language-speakers even beyond their ethnic homeland. Beyond self-reported identity salience, we find that these socioeconomic transformations resulted in different patterns of inter-ethnic marriage. Publishing contributed to the construction of more porous boundaries than cash crops, as evidenced by higher levels of exogamy with linguistically closely related ethnic outsiders.

In shedding light on some of the historical processes causing group formation and politicization, we highlight key underlying factors that may confound analyses of contemporary ethnic

²⁶The publications coefficient loses significance in Afrobarometer for ethnic-level results among ethnicities with history of publishing. Therefore, to address further concerns, Appendix Table A11 shows the generally insignificant and small effect of treatments when education and living standards are used as outcomes.

politics—such as contestation over land and cross-cutting languages.²⁷ These dynamics require greater attention among scholars of ethnic politics and conflict.²⁸

Our findings also have important implications for understanding the impact of colonialism on ethnicity. Much existing scholarship emphasizes the top-down effects of colonial social engineering and indirect rule on ethnic politicization (see for example Mamdani (1996) and Posner (2005) and more recently Ali et al. (2019) and McNamee (2019).) In contrast, our analysis demonstrates the importance of broader social and economic forces, which not only preceded colonialism but in many ways were key drivers of it. We further demonstrate that these processes led to striking within-country differences in boundary configurations—belying the notion that colonialism uniformly molded or “fixed” ethnic boundaries. Instead, identity (re)construction as much arose from the strategic actions of African farmers, landowners and elites as well as missionaries and their followers responding to opportunities and constraints brought about by economic and technological change in the 19th and 20th centuries.

²⁷On these points, see respectively, Boone (2014) and Laitin (2000a).

²⁸See Klaus (2020) for a recent such example.

References

- Adediran, Biodum. 1984. "Yoruba Ethnic Groups or a Yoruba Ethnic Group? A Review of the Problem of Ethnic Identification." *África* 7:57–70. (Cit. on p. 9).
- Adesoji, Abimbola O. 2017. "Yoruba ethnic nationalism, power elite politics and the Nigerian state, 1948–2007." *African Identities* 15 (2): 187–207. (Cit. on p. 10).
- Ahlerup, Pelle, and Ola Olsson. 2012. "The Roots of Ethnic Diversity." *Journal of Economic Growth* 17 (2): 71–102. (Cit. on p. 7).
- Ali, Merima, et al. 2019. "Colonial legacy, state-building and the salience of ethnicity in Sub-Saharan Africa." *The Economic Journal* 129 (619): 1048–1081. (Cit. on pp. 5, 7, 14, 28, 30).
- Anderson, Benedict O G. 2006. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso. (Cit. on pp. 3, 5, 7, 9, 21).
- Ashraf, Quamrul, and Oded Galor. 2013. "The 'Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review* 103 (1): 1–46. (Cit. on p. 7).
- Barth, Fredrik. 1969. *Ethnic Groups and Boundaries: The Social Organization of Culture Difference*. Allen / Unwin. (Cit. on pp. 3, 6).
- Bates, Robert H. 1983. "Modernization, ethnic competition, and the rationality of politics in contemporary Africa." In *State Versus Ethnic Claims: African Policy Dilemmas*, 152–171. Boulder, CO: Westview Press. (Cit. on pp. 2, 6, 7).
- Berghe, Pierre L. van den. 1981. *The Ethnic Phenomenon*. Elsevier. (Cit. on p. 7).
- Berman, Bruce J. 1998. "Ethnicity, Patronage and the African State: The Politics of Uncivil Nationalism." *African affairs* 97 (388): 305–341. (Cit. on p. 7).
- Berry, Sara. 1993. *No condition is permanent: The social dynamics of agrarian change in sub-Saharan Africa*. University of Wisconsin Pres. (Cit. on p. 8).
- Betz, Timm, Scott J. Cook, and Florian M. Hollenbach. 2019. "Spatial interdependence and instrumental variable models." *Political Science Research and Methods*: 1–16. (Cit. on p. 16).
- Birnir, Jóhanna K, et al. 2018. "Introducing the AMAR (All Minorities at Risk) Data." *Journal of Conflict Resolution* 62 (1): 203–226. (Cit. on p. 2).
- Birnir, Jóhanna K, et al. 2015. "Socially Relevant Ethnic Groups, Ethnic Structure, and AMAR." *Journal of Peace Research* 52 (1): 110–115. (Cit. on pp. 2, 14).

- Bockstette, Valerie, Areendam Chanda, and Louis Putterman. 2002. "States and Markets: The Advantage of an Early Start." *Journal of Economic Growth* 7 (4): 347–369. (Cit. on p. 7).
- Boni, Stefano. 2006. "Indigenous Blood and Foreign Labor: The 'Ancestralization' of Land Rights in Sefwi (Ghana)." Chap. 8 in *Land and the Politics of Belonging in West Africa*. Ed. by Carola Kuba Richard; Lentz. Brill. (Cit. on p. 8).
- Boone, Catherine. 2014. *Property and political order in Africa: Land rights and the structure of politics*. Cambridge University Press. (Cit. on pp. 8, 30).
- . 2017. "Sons of the soil conflict in Africa: institutional determinants of ethnic conflict over land." *World Development* 96:276–293. (Cit. on p. 8).
- Cagé, Julia, and Valeria Rueda. 2016. "The Long-Term Effects of the Printing Press in Sub-Saharan Africa." *American Economic Journal: Applied Economics* 8, no. 3 (): 1–31. (Cit. on p. 9).
- Caselli, Francesco, and Wilbur John Coleman. 2013. "On the theory of ethnic conflict." *Journal of the European Economic Association* 11 (1): 161–192. (Cit. on pp. 3, 8, 9, 27).
- Cervellati, Matteo, Giorgio Chiovelli, and Elena Esposito. 2018. "Bite and Divide: Ancestral Exposure to Malaria and the Emergence and Persistence of Ethnic Diversity in Africa." (Cit. on p. 21).
- Chandra, Kanchan. 2012. *Constructivist Theories of Ethnic Politics*. New York: Oxford University Press. (Cit. on p. 5).
- . 2006. "What is Ethnic Identity and Does It Matter?" *Annual Review of Political Science* 9:397–424. (Cit. on pp. 2, 6).
- . 2004. *Why Ethnic Parties Succeed: Patronage and Head Counts in India*. Cambridge: Cambridge University Press. (Cit. on p. 2).
- Chaney, Eric. 2016. "Religion and the rise and fall of Islamic science." Oxford. (Cit. on p. 12).
- Chimhundu, Herbert. 1992. "Early Missionaries and the Ethnolinguistic Factor during the 'Invention of Tribalism' in Zimbabwe." *The Journal of African History* 33 (1): 87–109. (Cit. on p. 9).
- Cordell, Dennis D, Joel W Gregory, and Victor Piché. 1996. *Hoe and Wage : A Social History of a Circular Migration System in West Africa*. Boulder, Colo.: Westview Press. (Cit. on p. 8).
- Eifert, Benn, Edward Miguel, and Daniel N Posner. 2010. "Political Competition and Ethnic Identification in Africa." *American Journal of Political Science* 54 (2): 494–510. (Cit. on p. 14).

- Englebert, Pierre, Stacy Tarango, and Matthew Carter. 2002. "Dismemberment and Suffocation: A Contribution to the Debate on African Boundaries." *Comparative Political Studies* 35 (10): 1093–1118. (Cit. on p. 7).
- Falola, Toyin. 1999. *Yoruba Gurus: Indigenous Production of Knowledge in Africa*. Africa World Press. (Cit. on p. 9).
- Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8 (2): 195–222. (Cit. on p. 2).
- . 2006. "Ethnic Mobilization and Ethnic Violence." In *The Oxford Handbook of Political Economy*, ed. by Donald A Wittman and Barry R Weingast, 852–868. Oxford: Oxford University Press. (Cit. on pp. 2, 6).
- . 1999. "Why ethnic politics and 'pork'tend to go together." In *an SSRC-MacArthur sponsored conference on "Ethnic Politics and Democratic Stability," University of Chicago*. Citeseer. (Cit. on pp. 6, 8).
- Frankema, Ewout, Jeffrey Williamson, and Pieter Woltjer. 2018. "An Economic Rationale for the West African Scramble? The Commercial Transition and the Commodity Price Boom of 1835–1885." *The Journal of Economic History* 78 (1): 231–267. (Cit. on p. 7).
- Gennaioli, Nicola, and Ilija Rainer. 2007. "The Modern Impact of Precolonial Centralization in Africa." *Journal of Economic Growth* 12 (3): 185–234. (Cit. on p. 7).
- Habyarimana, James, et al. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action*. New York: Russell Sage Foundation. (Cit. on p. 2).
- Hance, William A, Vincent Kotschar, and Richard J Peterec. 1961. "Source Areas of Export Production in Tropical Africa." *Geographical Review* 51 (4): 487–499. (Cit. on pp. 3, 11).
- Hechter, Michael. 1978. "Group formation and the cultural division of labor." *American Journal of Sociology* 84 (2): 293–318. (Cit. on p. 14).
- Hill, Polly. 1963. *The migrant cocoa-farmers of southern ghana*. Cambridge University Press. (Cit. on p. 8).
- Hirschman, Albert O. 1977. "A generalized linkage approach to development, with special reference to staples." *Economic development and cultural change* 25:67. (Cit. on p. 7).
- Hogendorn, Jan S. 1969. "Colonialism in Africa 1870-1960." Chap. Economic I in *Colonialism in Africa, 1870-1960: The Economics of Colonialism*. London: Cambridge University Press. (Cit. on p. 7).

- Hopkins, Anthony G. 1973. *An Economic History of West Africa*. New York: Columbia University Press. (Cit. on p. 7).
- Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. Berkeley: University of California Press. (Cit. on pp. 2, 7).
- Ichino, Nahomi, and Noah L Nathan. 2013. "Crossing the Line: Local Ethnic Geography and Voting in Ghana." *American Political Science Review* 107 (2): 344–361. (Cit. on p. 27).
- Jedwab, Remi, Felix zu Selhausen, and Alexander Moradi. 2018. *The economics of missionary expansion: evidence from Africa and implications for development*. Tech. rep. University of Sussex. (Cit. on pp. 4, 16).
- Kelly, Morgan. 2019. "The Standard Errors of Persistence." *UCD Working Papers Series DP 13783* (19/3). (Cit. on p. 25).
- Klaus, Kathleen. 2020. *Political Violence in Kenya: Land, Elections, and Claim-Making*. Cambridge University Press. (Cit. on p. 30).
- Laitin, David D. 2000a. "Language conflict and violence: the straw that strengthens the camel's back." *European Journal of Sociology/Archives Européennes de Sociologie* 41 (1): 97–137. (Cit. on p. 30).
- . 2007. *Language repertoires and state construction in Africa*. Cambridge University Press. (Cit. on p. 9).
- . 2000b. "What is a language community?" *American Journal of political science*: 142–155. (Cit. on pp. 4, 11).
- Laitin, David D, and Daniel Posner. 2001. "The Implications of Constructivism for Constructing Ethnic Fractionalization Indices." *APSA-CP: Newsletter of the Organized Section in Comparative Politics of the American Political Science Association* 12 (1): 13–17. (Cit. on p. 5).
- Lemarchand, Rene. 1980. "The Politics of Sara Ethnicity: A Note on the Origins of the Civil War in Chad (La politique d'ethnicité sara: note sur les origines de la guerre civile au Tchad)." *Cahiers d'Études africaines*: 449–471. (Cit. on p. 8).
- Lentz, Carola. 2013. *Land, Mobility, and Belonging in West Africa: natives and strangers*. Indiana University Press. (Cit. on p. 8).
- Mamdani, Mahmood. 1996. *Citizen and Subject: Contemporary Africa and the Legacy of Late Colonialism*. Princeton: Princeton University Press. (Cit. on pp. 7, 8, 30).

- Mann, Michael, and Valerie Sanders. 1994. *A bibliography of African language texts*. London: Hans Zell Publishers. (Cit. on pp. 3, 12, 13, 17, 23).
- McNamee, Lachlan. 2019. "Indirect colonial rule and the salience of ethnicity." *World Development* 122:142–156. (Cit. on pp. 5, 7, 30).
- Michalopoulos, Stelios. 2012. "The Origins of Ethnolinguistic Diversity." *The American Economic Review* 102 (4): 1508. (Cit. on p. 7).
- Michalopoulos, Stelios, and Elias Papaioannou. 2013. "Pre-colonial Ethnic Institutions and Contemporary African Development." *Econometrica* 81 (1): 113–152. (Cit. on p. 7).
- . 2016. "The Long-Run Effects of the Scramble for Africa." *American Economic Review* 106, no. 7 (7): 1802–1848. (Cit. on pp. 5, 7).
- Miguel, Edward. 2004. "Tribe or Nation? Nation Building and Public Goods in Kenya Versus Tanzania." *World politics* 56 (3): 327–362. (Cit. on p. 10).
- Müller-Crepon, Carl, Yannick Pengl, and Nils-Christian Bormann. 2020. "Linking Ethnic Data from Africa." Working Paper. (Cit. on p. 4).
- Murdock, George Peter. 1959. *Africa: Its Peoples and Their Culture History*. New York: McGraw-Hill. (Cit. on p. 11).
- . 1967. *Ethnographic Atlas: a Summary*. Pittsburg University Press. (Cit. on pp. 11, 28).
- Nettle, Daniel. 1998. "Explaining Global Patterns of Language Diversity." *Journal of anthropological archaeology* 17 (4): 354–374. (Cit. on p. 7).
- Nunn, Nathan. 2008. "The Long-term Effects of Africa's Slave Trades." *The Quarterly Journal of Economics* 123 (1): 139–176. (Cit. on p. 7).
- Parkin, F. 1979. *Marxism and Class Theory: A Bourgeois Critique*. Columbia University Press. (Cit. on pp. 3, 7, 8).
- Peel, John David Yeadon. 2003. *Religious Encounter and the Making of the Yoruba*. Indiana University Press. (Cit. on p. 9).
- Posner, Daniel N. 2005. *Institutions and Ethnic Politics in Africa*. New York: Cambridge University Press. (Cit. on pp. 5, 7, 30).
- . 2004a. "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science* 48 (4): 849–863. (Cit. on pp. 3, 14).
- . 2003. "The Colonial Origins of Ethnic Cleavages: The Case of Linguistic Divisions in Zambia." *Comparative Politics*: 127–146. (Cit. on pp. 3, 9).

- . 2004b. “The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi.” *American Political Science Review* 98 (4): 529–545. (Cit. on p. 7).
- Ranger, Terrence. 1989. “Missionaries, Migrants, and the Manyika: The Invention of Ethnicity in Zimbabwe.” Chap. 4 in *The Creation of Tribalism in Southern Africa*. Berkeley: Univ of California Press. (Cit. on pp. 3, 9).
- Robinson, Amanda Lea. 2017. “Ethnic diversity, segregation and ethnocentric trust in Africa.” *British Journal of Political Science*: 1–23. (Cit. on p. 15).
- . 2014. “National versus Ethnic Identification in Africa: Modernization, Colonial Legacy, and the Origins of Territorial Nationalism.” *World Politics* 66 (4): 709–746. (Cit. on p. 14).
- Rodney, Walter. 1972. *How Europe Underdeveloped Africa*. London: Bogle-L'Ouverture Publications. (Cit. on p. 7).
- Roessler, Philip. 2016. *Ethnic Politics and State Power in Africa*. New York: Cambridge University Press. (Cit. on p. 2).
- Rowling, Frank, and Charles-Edward Wilson. 1923. *Bibliography of African Christian literature*. London: Conference of Missionary Societies of Great Britain / Ireland. (Cit. on pp. 3, 12, 13, 15, 17).
- Sasaki, Yu. 2017. “Publishing Nations: Technology Acquisition and Language Standardization for European Ethnic Groups.” *The Journal of Economic History* 77 (4): 1007–1047. (Cit. on pp. 5, 21).
- Simons, Gary F, and Charles D Fennig. 2017. *Ethnologue: Languages of the World*. Dallas, TX: SIL. (Cit. on p. 11).
- Strommer, Martina A. 2015. “Imagined Communities, Invented Tribe? Early Missionary Language Documentation and the Creation of the Herero.” In *Colonialism and Missionary Linguistics*, ed. by Klaus Zimmerman and Birte Kellermeier-Rehbein, 102–128. Colonial and Postcolonial linguistics. Berlin: De Gruyter. (Cit. on p. 9).
- Vogt, Manuel, et al. 2015. “Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family.” *Journal of Conflict Resolution* 59 (7): 1327–1342. (Cit. on pp. 2, 3, 14).
- Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*. Vol. 1. Univ of California Press. (Cit. on pp. 3, 5, 7, 8).

- Wimmer, Andreas. 2013. *Ethnic Boundary Making: Institutions, Power, Networks*. Oxford University Press. (Cit. on pp. 3, 6).
- . 2016. “Is Diversity Detrimental? Ethnic Fractionalization, Public Goods Provision, and the Historical Legacies of Stateness.” *Comparative Political Studies* 49 (11): 1407–1445. (Cit. on p. 7).
- Woodberry, Robert D. 2012. “The Missionary Roots of Liberal Democracy.” *American Political Science Review* 106 (02): 244–274. (Cit. on p. 9).

Supplementary Materials (Online)

Tables	A1
Pre-registered Analyses	A1
Supplementary, non-preregistered Analyses	A3
Data	A4
Afrobarometer	A5
Figures	A13
Bloc-Voting	A18
Additional DHS results: Geographic Persistence	A19
Additional DHS results: Cultural Persistence	A22
Additional DHS results: Robustness	A23
Ethnic Conflict	A27

Tables

Pre-registered Analyses

Table A1: Preregistered Analyses

Pre-Registered Analysis	Implementation
<i>Ethnic politicization and salience, group-level</i>	
Group-level politicization Group-level ethnic politicization based on a group listed as relevant in EPR and PREG under the same or a clearly synonymous name.	As described. See Figure 2.
Group-level salience Group-level ethnic salience based on a group listed as socially-relevant in AMAR.	As described. See Figure A6.
<i>Ethnic politicization, individual-level</i>	
Ethnic vs national identity Ethnic salience using Afrobarometer survey responses to whether individuals identify more in ethnic or national terms -Geographic-matching: survey locations matched to Ethnologue polygons -Ethnic-matching: based on language individuals speak at home	As described. See Table 1. As described. See Table 2. As a robustness, we also tested matching on declared ethnicity (which has fewer observations) and results remain unchanged.
Bloc-voting Homogeneity of political preferences among members of the same ethnic group.	Implemented with vote intention. See Figure A7.
Mechanisms Mechanisms analysis: Ethnic salience among stayers vs leavers - Geographic-matching: sub-group analysis of leavers - Ethnic-matching: sub-group analysis of leavers	As described. See Table 1, Column 5. As described. See Table 2, Columns 5-6.

Table A2: Preregistered Analyses - Continuation

Pre-Registered Analysis	Continuation	Implementation
<i>Ethnic boundary-making</i>		
Inter-ethnic marriage		
Inter-ethnic marriage using data from the Demographic and Health Surveys at all 16 levels of Ethnologue’s linguistic hierarchy		
- Geographic-matching: Assignment based on survey location		As described. See Figure 3
- Ethnic-matching: Assignment based on husband’s ethnicity		As described. See Figure 4
- Ethnic-matching: Assignment based on wife’s ethnicity		As described. See Figure A11 .
Stayers vs leavers analysis of inter-ethnic marriage		
- Geographic-matching: sub-group analysis of stayers and leavers		As described. See Figure A10.
- Ethnic-matching: sub-group analysis of leavers, assignment based on husband or wife’s ethnicity.		As described. See Figures 4 and A11
 Ethnocentric Trust		
Ethnocentric trust: based on two questions from Afrobarometer Round 3 about trust in co-ethnics and non-co-ethnics		Deviation: Given the high correlation between answers to trust coethnics and trust non-coethnics, ethnocentric trust is computed instead as a binary variable flagging those reporting high coethnic trust but low <i>general trust</i> (as opposed to low non-coethnic trust).
- Geographic-matching		See Table A5, Columns (1)-(3).
- Ethnic-matching		See Table A5, Columns (4)-(6).
Stayers vs leavers analysis of ethnocentric trust		Not pursued given weakness in trust data described above and inconclusive results of baseline ethnocentric trust analysis reported in Table A5.
 Ethnic Conflict		
Ethnic conflict: aggregate all clearly ethnicity-related events from ACLED to the Ethnologue polygons		Use ACLED identity militia measure. See Figure A17.
<i>Addressing endogeneity</i>		
Instrumental Variables		
Use cash crop suitability to run 2SLS-IV models to identify the causal effects of colonial cash crop extraction on ethnic identities.		As described. Lines 3-4 in Figure 2. Column (4) in Table 1. Column 5 in Table 2. Figure 5.
Intensive Margin within exposed areas		
Intensive margin: subsetting analysis to groups with at least one publication and estimating the effect of magnitude in publication record		As described. See Tables 1 and 2, and Figure 6.
Linguistic proximity		
Analysis of linguistically-proximate groups to test spillovers		Not implemented yet. Spillover analysis is an avenue for future research.

Supplementary, non-preregistered Analyses

Table A3: Non-pre-registered Analyses

Non-pre-registered Analysis	Results
Additional ethnic conflict specifications Two ethnic conflict outcomes from the SCAD database (Salehyan et al. 2012) as well as non-state conflict events from UCDP-GED. See discussion on page A27	See Figure A17.
Alternative potential channels of ethnicization and robustness checks Proximate effects of education, urbanization, and wealth.	See Table A11.
Pre-colonial political centralization	See Table A12 & Figure A16.
British colonial legacy	See Table A13 & Figure A15.
Heterogeneity by local ethnic diversity	See Figure A13.
Controlling for effects of religion	See Figure A14.

Data

Table A4: Data Sources

(1)	(2) Name	(3) Purpose	(4) Match to Ethnologue
<i>Unit of analysis</i>	Ethnologue	Unit of analysis	–
<i>Print technologies</i>	Cagé and Rueda (2016)	Treatment	Spatial
	Rowling and Wilson (1923)	Treatment	Ethnic
	Mann and Sanders (1994)	Treatment	Ethnic
<i>Cash crops</i>	Hance, Kotschar, and Peterec (1961)	Treatment	Spatial
<i>Ethnic characteristics</i>	Afrobarometer	Outcomes (ethnic vs. national identity salience, inter-ethnic trust and bloc-voting)	Ethnic & Spatial
	DHS	Outcomes (inter-ethnic marriage)	Ethnic & Spatial
	EPR & PREG	Outcomes (politically salient groups)	Ethnic
	Murdock (1959)	Controls (historical)	Ethnic and Spatial

Notes: This table summarizes the data sources used for this analysis. Column (1) states the type of data, (2) cites the sources, (3) states the purpose, and (4) the type of match required to merge to Ethnologue. A “Spatial” merge refers to a merge to Ethnologue based on the spatial correspondence between geographic information in data (2) and the ethnic homelands mapped in Ethnologue. An “Ethnic” merge refers to a merge between the ethnolinguistic group in (2) and in its linked ethnolinguistic name in Ethnologue.

Afrobarometer

Trust

In the registered report, we conceived that another observable implication of the effects of cash crops and print technologies on boundary-making was via co-ethnic trust. In the spirit of Nunn and Wantchekon (2011), we look at three binary outcomes of interest that flag whether respondents: (i) declare high level of trust towards coethnics, (ii) declare high level of trust towards most people, and (iii) have a coethnic-trust premium (i.e., report trusting coethnics, but not trusting people in general.) Results are reported below in Table A5. We observe that at both geographic and ethnic-levels, the cash crop treatments are associated with lower levels of coethnic trust (Columns (1) and (4)), but this appears to be a function of lower levels of trust in general (Columns (2) and (5)). Indeed, the cash crop treatment is associated with a positive coethnic trust premium (only significant in ethnic-level specification, in Column (6)). No robust pattern appears for print technologies. These results align with the hypothesis that historical cash crop production engendered more exclusionary identities. However, as these outcomes are only collected in Round (3) of the Afrobarometer, they should be interpreted with caution. Better data, perhaps collected from behavioral field experiments, would be valuable in further testing this hypothesis.

Table A5: Cash Crops, Print Technologies, and Ethnic Trust

	Trust					
	Geographic-Level			Ethnic-Level		
	(1)	(2)	(3)	(4)	(5)	(6)
Cash crops USD plm2	-0.020** (0.009)	-0.020*** (0.005)	0.003 (0.009)			
Pubs pth pop (1923)	-0.011* (0.006)	-0.004 (0.005)	0.004 (0.009)			
Cash crops USD plm2				-0.061*** (0.016)	-0.017* (0.010)	0.037** (0.016)
Pubs pth pop (1923)				0.002 (0.008)	-0.003 (0.006)	-0.009 (0.011)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	No	No	No	No
Fixed Effect	Country-Round	Country-Round	Country-Round	Town	Town	Town
Ethnic Stayer/Leaver	Both	Both	Both	Both	Both	Both
Mean dep. var.	0.255	0.182	0.905	0.255	0.182	0.905
Observations	21,883	73,227	21,449	20,538	69,696	20,126
R ²	0.127	0.069	0.048	0.276	0.233	0.203

Notes: $p < 0.1$:*, $p < 0.05$ **:*, $p < 0.01$:***. The table reports standardised OLS estimates. Standard errors are reported in parenthesis and clustered at the location level. Columns (1)-(3) use specification at the geographic level, like in Table 1 in the paper, whereas Columns (4)-(6) use specification at the ethnic level, like in Table 2 of the paper. The outcome in Columns (1) and (4) is a binary variable equal to one if respondent has high level of trust towards coethnics, Columns (2) and (5) is a binary variable equal to one if respondent has high level of trust towards most people, and in Columns (2) and (5) it is equal to one if the declared level of trust towards co-ethnics is larger than the declared general level of trust towards most people.

Robustness

Table A6: Cash Crops, Print Technologies, and Political Relevance - Intensive Margin - Exclusive Links

	EPR and PREG Political Relevance (Y/N)					
	OLS		Language in Bibliographies			
	(1)	(2)	(3)	(4)	(5)	(6)
Cash crops pkm2 USD	0.067*** (0.013)	0.091** (0.041)	0.233*** (0.054)	0.189 (0.129)		
Pubs pth pop	0.086 (0.060)	0.120* (0.060)	0.082 (0.056)	0.110** (0.052)	0.108* (0.059)	0.130** (0.052)
Cash crop suitability					0.063* (0.032)	0.046 (0.050)
Geography controls	yes	yes	yes	yes	yes	yes
Historical controls	yes	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes	yes
Mean dep. var.	0.035	0.017	0.12	0.073	0.12	0.073
Sd dep. var.	0.185	0.128	0.326	0.26	0.326	0.26
Observations	599	594	178	175	178	175
R ²	0.227	0.351	0.347	0.504	0.323	0.494

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports OLS estimates and IV. Standard errors are reported in parenthesis and clustered at the country level. The dependent variable is a binary variable equal to one if the ethnic group is listed as politically relevant in EPR (columns (1), (3), and (5)), or PREG (columns (2), (4), and (6)). Only ethnic groups that are exclusively listed (1-to-1 match) in EPR (resp. PREG) will have an outcome equal to 1. Columns (3) to (6) restrict the sample to ethnic groups with languages mentioned in publication records.

Table A7: Cash Crops, Print Technologies, and Political Relevance - Intensive Margin - All Links

	EPR and PREG Political Relevance (Y/N)					
	OLS		Language in Bibliographies			
	(1)	(2)	(3)	(4)	(5)	(6)
Cash crops pkm2 USD	0.070 (0.065)	0.091** (0.041)	0.165 (0.149)	0.189 (0.129)		
Pubs pth pop	0.012 (0.063)	0.120* (0.060)	0.052 (0.075)	0.110** (0.052)	0.094 (0.087)	0.130** (0.052)
Cash crop suitability					0.124** (0.047)	0.046 (0.050)
Geography controls	yes	yes	yes	yes	yes	yes
Historical controls	yes	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes	yes
Mean dep. var.	0.293	0.133	0.407	0.238	0.407	0.238
Sd dep. var.	0.455	0.34	0.492	0.426	0.492	0.426
Observations	599	594	178	175	178	175
R ²	0.481	0.351	0.515	0.504	0.522	0.494

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports OLS estimates and IV. Standard errors are reported in parenthesis and clustered at the country level. The dependent variable is a binary variable equal to one if the ethnic group is listed as politically relevant in EPR (columns (1), (3), and (5)), or PREG (columns (2), (4), and (6)). All ethnic groups that are listed (many-to-1 match) in EPR (resp. PREG) will have an outcome equal to 1. Columns (3) to (6) restrict the sample to ethnic groups with languages mentioned in publication records.

Table A8: Geographical Persistence in Ethnic Identity - Other treatment

	Geographic- level - Ethnic vs National Id				
	(1)	(2)	(3)	(4)	(5)
Cash crops USD pkm2	0.013*** (0.004)		0.012*** (0.004)		-0.015** (0.007)
Cash Crop Suitability				-0.007 (0.006)	
Pubs pth today pop SOAS		-0.001 (0.004)	-0.001 (0.004)	-0.002 (0.004)	0.002 (0.006)
Individual controls	yes	yes	yes	yes	yes
Historical and Geo controls	yes	yes	yes	yes	yes
FE	Country-Round	Country-Round	Country-Round	Country-Round	Country-Round
Ethnic Stayer/Leaver	Both	Both	Both	Both	Leaver
Mean dep. var.	0.1294	0.1294	0.1294	0.1294	0.1294
Observations	136,255	132,273	132,273	132,273	44,759
R ²	0.041	0.041	0.042	0.041	0.044

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports standardised OLS estimates (“beta coefficients”). Standard errors are reported in parenthesis and clustered at the language spoken level. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The treatment is the number of publications listed in Mann and Sanders (1994), normalised by a current estimate of population speaking the language, according to JP.

Table A9: Cultural Persistence in Ethnic Identity - Other Treatment

	Ethnic-level - Ethnic vs National Id				In Biblio
	(1)	(2)	(3)	(4)	
Cash crops USD pkm2	-0.010 (0.007)		-0.011 (0.007)	-0.005 (0.010)	0.018 (0.016)
Pubs pth today pop SOAS		0.008* (0.005)	0.009* (0.005)	0.023*** (0.007)	0.029*** (0.009)
Individual controls	Yes	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	No	No	No
Fixed Effect	Town	Town	Town	Town	Town
Ethnic Stayer/Leaver	Both	Both	Both	Leaver	Leaver
Mean dep. var.	0.1294	0.1294	0.1294	0.12	0.11
Observations	127,000	130,114	126,911	45,322	31,569
R ²	0.202	0.200	0.202	0.263	0.269

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports standardised OLS estimates (“beta coefficients”). Standard errors are reported in parenthesis and clustered at the location level. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The treatment is the number of publications listed in Mann and Sanders (1994), normalised by a current estimate of population speaking the language, according to JP.

Table A10: Individual-level ethnic politicization- Bad Controls

	Geographic level		Ethnic level	
	(1)	(2)	(3)	(4)
Cash crops USD pkm2	0.010** (0.004)	0.010** (0.005)		
Pubs pth pop (1923)	0.038*** (0.006)	0.046*** (0.007)		
Cash crops USD pkm2			-0.015** (0.007)	-0.002 (0.014)
Pubs pth pop (1923)			0.010* (0.005)	0.007 (0.007)
Individual controls	Yes	Yes	Yes	Yes
Historical and Geo controls	Yes	Yes	No	No
Fixed Effect	Country-Round	Country-Round	Town	Town
Ethnic Stayer/Leaver	Both	Both	Both	Both
Sample	All	In Biblio	All	In Biblio
Mean dep. var.	0.129	0.13	0.129	0.13
Observations	133,380	89,702	127,413	73,106
R ²	0.036	0.042	0.199	0.207

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports standardised OLS estimates (“beta coefficients”). Standard errors are reported in parenthesis and clustered at the location level. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The regressions exclude potential “bad controls” capturing living standards, urbanization, and education.

Table A11: Alternative explanations: education and living standards as outcomes

	Geographic level - Education		Geographic level - Poverty		Ethnic level - Education		Ethnic level - Poverty	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cash crops USD pkm2	-0.004 (0.004)	-0.012*** (0.004)	-0.004 (0.004)	-0.004 (0.005)				
Pubs pth pop (1923)	0.001 (0.005)	-0.005 (0.005)	0.006 (0.004)	0.005 (0.005)				
Cash crops USD pkm2					0.012 (0.007)	-0.004 (0.012)	0.003 (0.006)	-0.006 (0.009)
Pubs pth pop (1923)					-0.003 (0.007)	-0.003 (0.008)	0.006 (0.004)	0.011** (0.005)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Historical and Geo controls	Yes	Yes	Yes	Yes	No	No	No	No
Fixed Effect	Country-Round	Country-Round	Country-Round	Country-Round	Town	Town	Town	Town
Ethnic Stayers/Leaver	Both	Both	Both	Both	Both	Both	Both	Both
Sample	All	In Biblio	All	In Biblio	All	In Biblio	All	In Biblio
Mean dep. var.	0.126	0.12	0.029	0.032	0.126	0.12	0.029	0.032
Observations	176,144	111,487	176,248	111,519	170,411	118,428	170,504	118,467
R ²	0.097	0.094	0.032	0.035	0.258	0.263	0.176	0.191

Notes: $p < 0.1$; * , $p < 0.05$; ** , $p < 0.01$; *** . The table reports standardized OLS estimates ("beta coefficients). Standard errors are reported in parenthesis and clustered at the location level. Outcome are all binary variables equal to one under the following conditions. Columns (1),(2),(5), and (6): respondent has higher education; Columns (3),(4), (7) and (8): respondent reports having gone without food many times or more during the year. The treatment is defined at the geographic level in Columns (1) to (4), and at the ethnic level in Columns (5) to (8). Columns (2), (4), (6), and (8) restrict the analysis to locations listed in bibliographic records.

Table A12: Pre-Colonial Centralisation

	Geographic-level		Ethnic-level	
	(1)	(2)	(3)	(4)
Cash crops USD pkm2	0.012*** (0.005)	0.011** (0.005)		
Pubs pth pop (1923)	0.040*** (0.006)	0.041*** (0.007)		
Murdock Centralisation	-0.020*** (0.006)	-0.026*** (0.008)		
Cash crops USD pkm2			-0.019*** (0.008)	-0.009 (0.011)
Pubs pth pop (1923)			0.012** (0.005)	0.012* (0.007)
Murdock Centralisation			-0.001 (0.007)	-0.002 (0.009)
Individual controls	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	No	No
Fixed Effect	Country-Round	Country-Round	Country-Round	Country-Round
Ethnic Stayer/Leaver	Both	Both	Both	Both
Sample	All	In Biblio	All	In Biblio
Mean dep. var.	0.129	0.13	NA	NA
Observations	110,628	84,325	110,615	91,641
R ²	0.039	0.043	0.198	0.207

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports OLS estimates. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The treatments, including Murdock centralisation, are defined at the location level (Columns (1) and (2)), and at the ethnic level (Columns (3) and (4)). Standard errors are reported in parenthesis and clustered at the location level.

Table A13: Cultural Persistence in Ethnic Identity - Colonial Administration Legacy

	Geographic-level		Ethnic-level	
	(1)	(2)	(3)	(4)
Cash crops USD pkm2	0.024*** (0.009)	0.034*** (0.012)		
Cash crops USD pkm2 x British Colony	-0.016 (0.010)	-0.024* (0.013)		
Pubs pth pop (1923)	0.022*** (0.006)	0.033*** (0.007)		
Pubs pth pop (1923) x British Colony	0.030*** (0.006)	0.024*** (0.007)		
Cash crops USD pkm2			-0.019* (0.010)	-0.033** (0.015)
Cash crops USD pkm2 x British Colony			0.019 (0.015)	0.021 (0.019)
Pubs pth pop (1923)			0.012* (0.007)	0.013 (0.008)
Pubs pth pop (1923) x British Colony			-0.005 (0.010)	-0.005 (0.012)
Individual controls	Yes	Yes	Yes	Yes
Historical and Geo controls	No	No	Yes	Yes
Fixed Effect	Country-Round	Country-Round	Town	Town
Sample	All	In Biblio	All	In Biblio
Mean dep. var.	0.1294	0.13	0.1294	0.13
Observations	125,390	86,838	125,486	85,014
R ²	0.043	0.047	0.201	0.192

Notes: $p < 0.1$:*, $p < 0.05$:**, $p < 0.01$:***. The table reports OLS estimates. Coefficients are standardised (“beta” coefficients). Standard errors are reported in parenthesis and clustered at the location level. The dependent variable is a binary variable equal to one if respondent declares a stronger ethnic than national identity. The treatments are defined at the location level (Columns (1) and (2)), and at the ethnic level (Columns (3) and (4))

DHS: Spatial Models

Table A14: Impacts from spatial models: Group-level political relevance

Outcome	Cash Crop Coef.	Average Direct Effect
EPR Link	0.1559	0.1600
EPR Excl. Link	0.0392	0.0287
PREG Link	0.2086	0.2227
PREG Excl. Link	0.0381	0.0465
AMAR Link	0.4262	0.4310
AMAR Excl. Link	0.3712	0.3773

Notes: The table reports raw coefficients and average direct effects from the group-level spatial IV models in Figure 2.

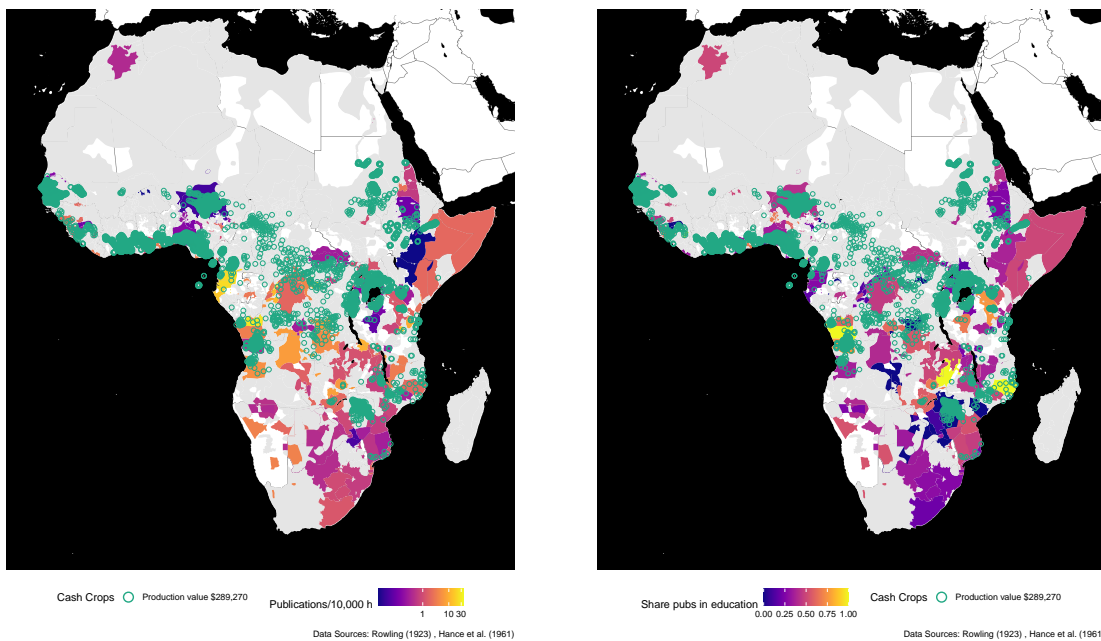
Table A15: Impacts from spatial models: Inter-ethnic marriages

Outcome	Crop Coef. (SAR)	Avg. Dir. Effect (SAR)	Crop Coef. (S2SLS)	Avg. Dir. Effect (S2SLS)
Exogamy L1	-0.0109	-0.0109	-0.0217	-0.0217
Exogamy L2	-0.0139	-0.0140	-0.0267	-0.0267
Exogamy L3	-0.0158	-0.0159	-0.0270	-0.0270
Exogamy L4	-0.0155	-0.0155	-0.0237	-0.0238
Exogamy L5	-0.0152	-0.0152	-0.0214	-0.0214
Exogamy L6	-0.0141	-0.0141	-0.0202	-0.0203
Exogamy L7	-0.0190	-0.0190	-0.0273	-0.0274
Exogamy L8	-0.0185	-0.0186	-0.0254	-0.0254
Exogamy L9	-0.0165	-0.0165	-0.0255	-0.0255
Exogamy L10	-0.0132	-0.0132	-0.0209	-0.0208
Exogamy L11-14	-0.0128	-0.0128	-0.0191	-0.0191
Exogamy L15	-0.0109	-0.0109	-0.0188	-0.0188
Exogamy L16	-0.0113	-0.0113	-0.0200	-0.0200

Notes: The table reports raw coefficients and average direct effects from the SAR and S2SLS-IV models of inter-ethnic marriage in Figure 5.

Figures

Figure A1: Description of Publications in Rowling and Wilson (1923)



(a) Publications per 10,000 inhabitants

(b) Share of education publications

Notes: Language homelands are mapped according to Ethnologue. Greyed regions are those for which there is no record of publications. Publications per 10,000 inhabitants in map (a) are computed using population estimates in Rowling and Wilson (1923). The share of publications in education are computed counting the numbers listed as education materials in Rowling and Wilson (1923). Each green circle locates 289,270 USD (1957) cash-crop export value.

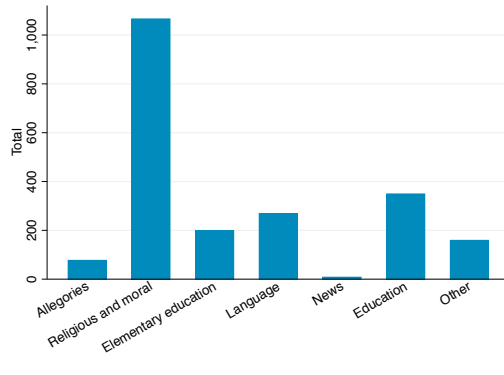
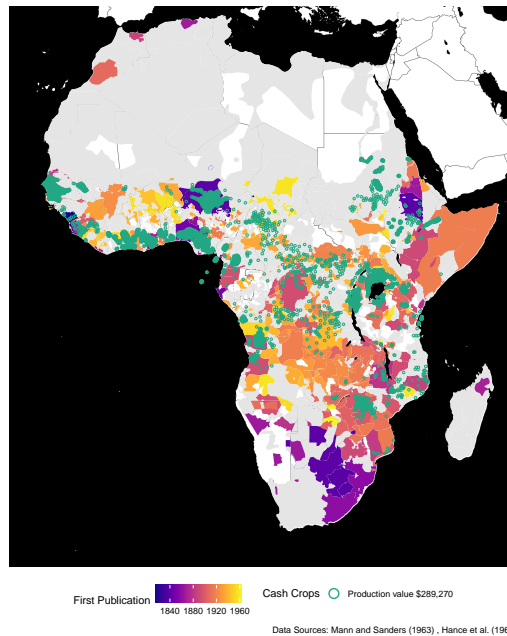


Figure A2: Christian publications in African languages up to 1923

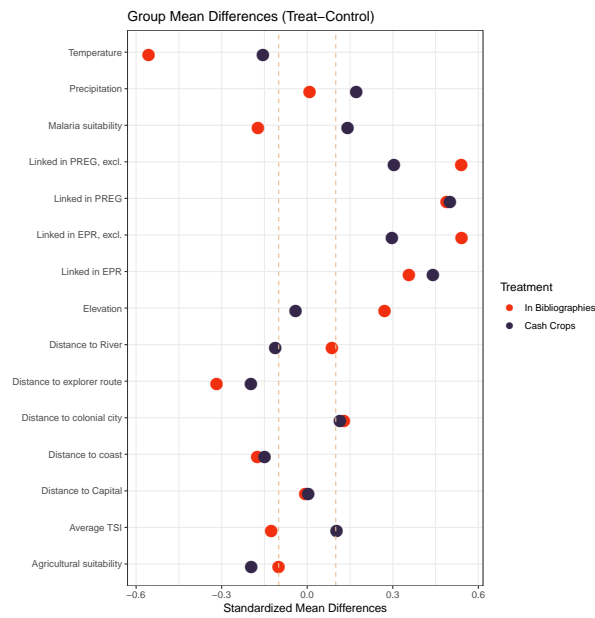
Notes: This chart gives the total number of Christian publications in African languages by type up to 1923, as described in Rowling and Wilson (1923)

Figure A3: Publications and Cash Crop locations in Mann and Sanders (1994)



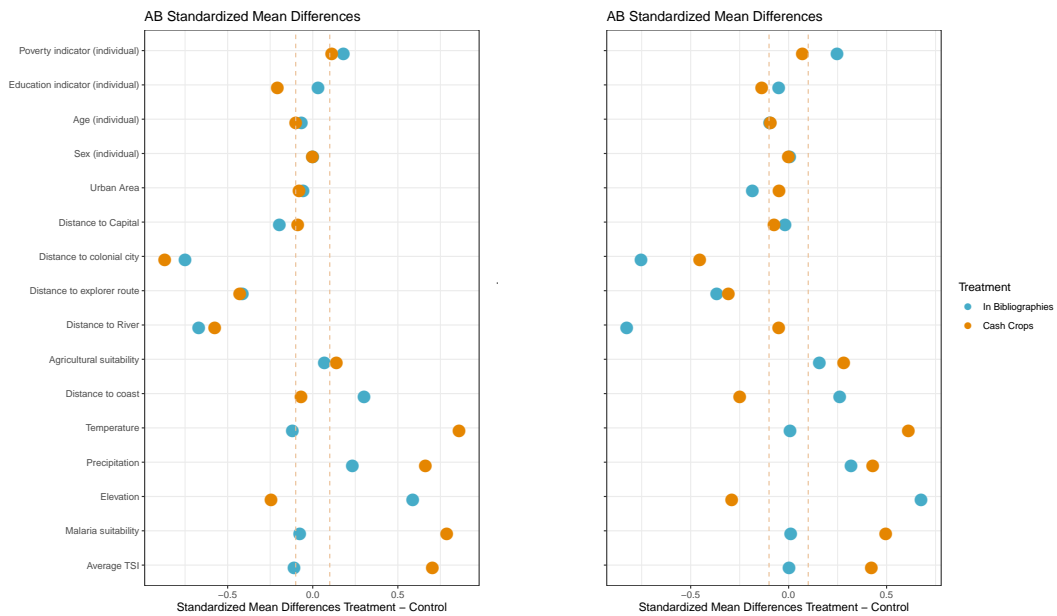
Notes: Language homelands are mapped according to Ethnologue. Greyed regions are Ethnologue polygons for which there is no record of publications. The colors indicated the number of publications listed in Mann and Sanders (1994). Each green circle locates 289,270 USD (1957) cash-crop export value.

Figure A4: Balance of Covariates for Group-Level Data



Notes: The figure shows the standardized mean differences in covariates in the group-level analysis. In each graph, dots of different colors represent the different treatments used to compare means: print technology (group listed in bibliographies), or exposure to cash crop technology. The dashed vertical lines represent the 10% confidence interval around 0.

Figure A5: Balance of Covariates for Afrobarometer

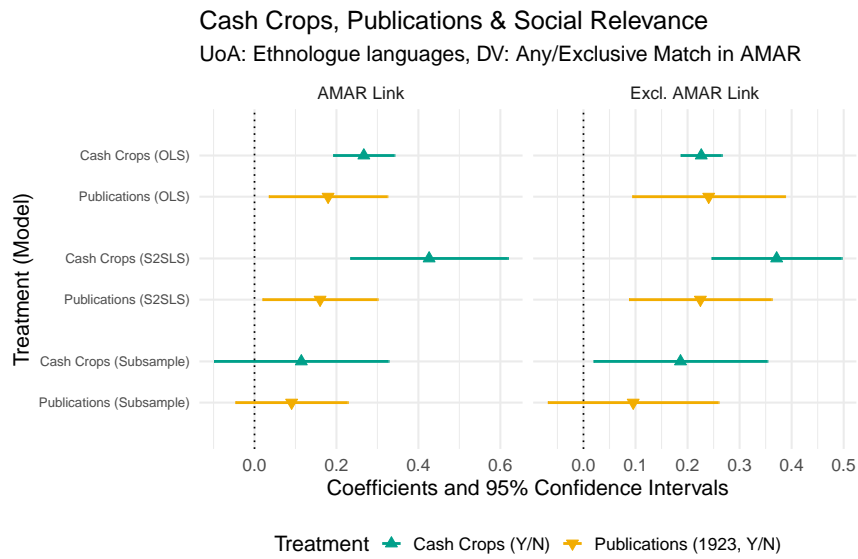


(a) Geographic level treatment

(b) Ethnic-level treatment

Notes: The figure shows the standardized mean differences in covariates in Afrobarometer. In each graph, dots of different colors represent the different treatments used to compare means: print technology (group listed in bibliographies), or exposure to cash crop technology. The dashed vertical lines represent the 10% confidence interval around 0.

Figure A6: Cash Crops, Print Technologies, and Political Relevance - AMAR

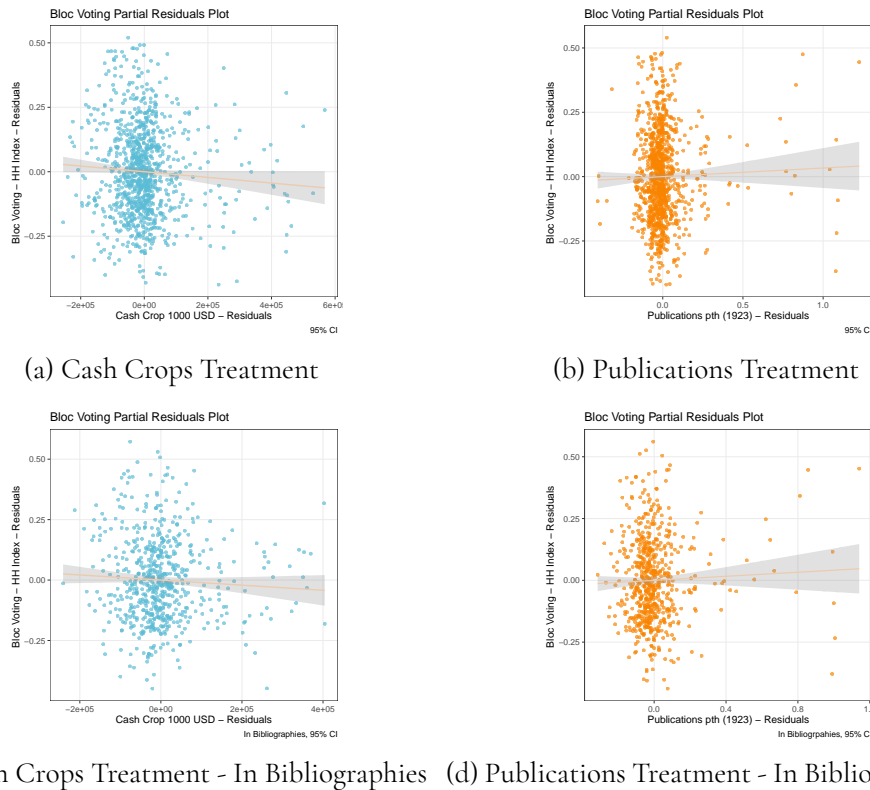


Notes: These figures summarise the results of 6 regression models. Each column gives the results of 2 different outcomes, which are all binary variables. “AMAR Link” flags whether an Ethnologue group is non-exclusively matched to a group listed as relevant in AMAR. “Excl AMAR Link” flags an exclusive (1-to-1) match between the same databases. Different treatment specifications are shown in lines. The first two lines report OLS estimates using binary treatments indicating whether Ethnologue groups were exposed to cash crop production and/or print technologies. In lines 3-4, cash crops are instrumented with the mean agro-climatic suitability for the five most important export crops using the S2SLS described in the text. In lines 5-6, the sample is restricted to Ethnologue polygons that experienced missionary activity.

Bloc-Voting

In the pre-registered plan, we also promised to look at bloc-voting as a measure of ethnic politicisation. To do so, we run a regression similar to those described in this section, using Afrobarometer. The outcome of interest is the Herfindahl Index of political party choices in a hypothetical voting scenario. In Afrobarometer, respondents are asked to choose the party for which they would vote if elections were to be held tomorrow. The results are reported in the online Appendix, in Figure A7. We only observe a positive association of publication treatments, robust to restricting the analysis to ethnicities listed in the Bibliographies. However the effect is not statistically significant. Further research, maybe using surveys that sampled so as to be representative of each ethnicity, is required to study the issue more carefully.

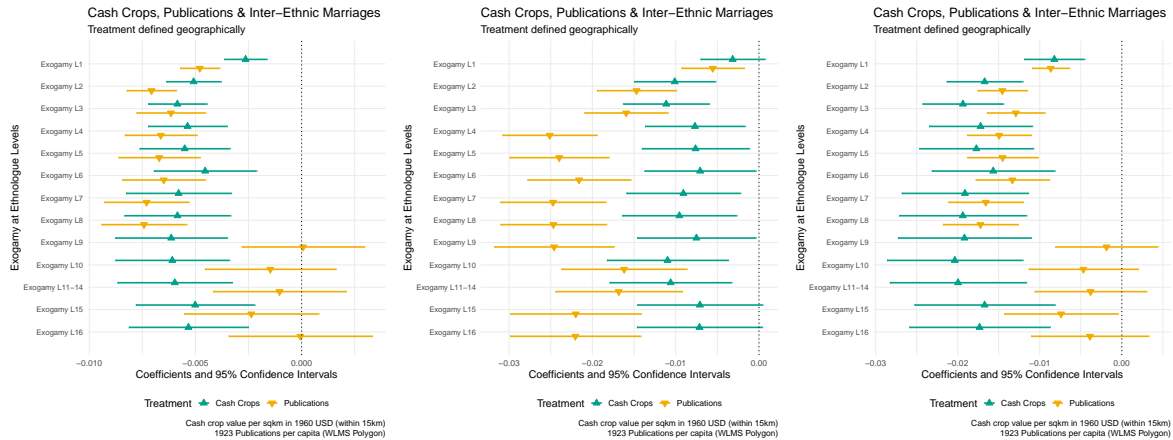
Figure A7: Bloc Voting in Afrobarometer



Notes: The figures show partial residuals correlations of the regression model described in section 4. The unit of observation is the ethnicity. The outcome of interest is the Herfindahl Index of political party choices in a hypothetical voting scenario. In Afrobarometer, respondents are asked to choose the party for which they would vote if elections were to be held tomorrow. A larger Herfindahl index indicates more homogeneity in the choice of political parties within the ethnic group. The slope of the linear fits in the plots gives OLS estimate of the treatment (cash crops in Figures A7a and A7c and print technologies in Figures A7b and A7d). The regression model includes historical and geographic controls described in the text, as well as country and survey round fixed effects. Figures A7a and A7b are from the regression with the entire sample, whereas Figures A7c and A7d are from regressions that restrict the analysis to ethnicity with languages listed in our bibliographies.

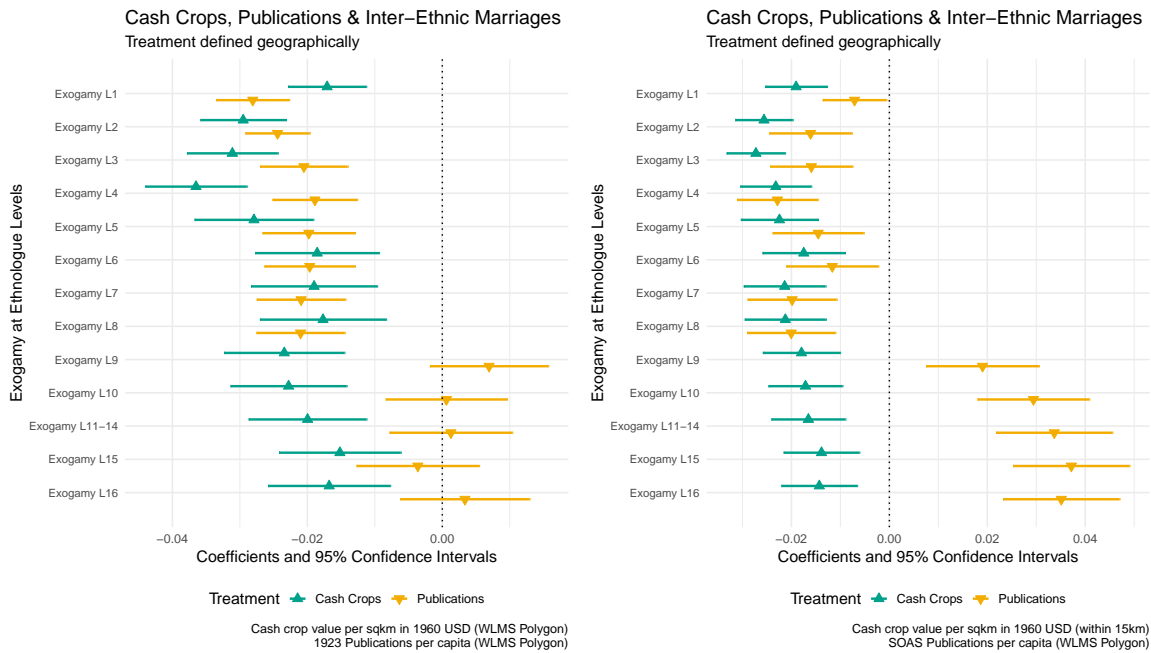
Additional DHS results: Geographic Persistence

Figure A8: Geographical persistence - unstandardized outcomes & substantive effect sizes



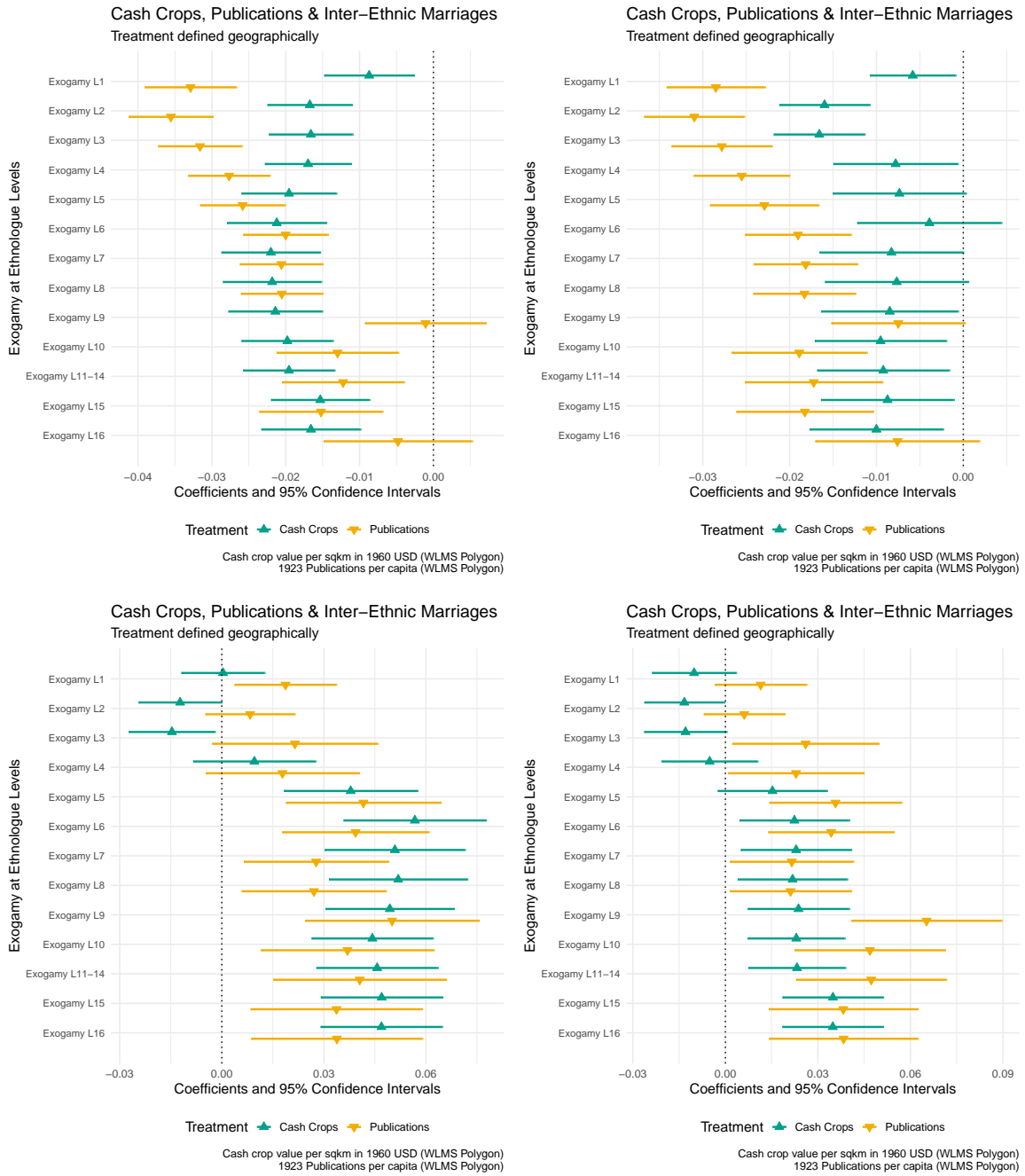
Notes: Each panel reports coefficients from 16 OLS models with unstandardized exogamy dummies along the Ethnologue language tree as dependent variables. Fixed effects, control variables, and standard errors remain equivalent to the baseline DHS geographic persistence models. The left-hand panel indicates that a one-standard deviation increase in historical cash crop production is associated with a 0.26-0.58 percentage points decrease in the probability of inter-ethnic marriage at the respective Ethnologue level. These effects amount to 3.24-10.16% reductions compared to the sample mean of the respective dependent variable. A one-standard deviation increase in publications per capita is associated with a 0.47 to 0.74 pp lower likelihood of inter-ethnic marriages at Ethnologue levels 1-8 which corresponds to 7.22-18.46% of the DV means. The middle panel reports extensive margins estimates which are larger in both absolute size and relative to the DV means than the standard deviation effects discussed above (Cash Crops: 0.32-1.11 pp, 6.01-19.46% from the mean; Publications: 0.56-2.51 pp, 10.96-34.71% from the mean). The right-hand panel replaces standardized linear predictors with logged versions of cash crop production value per 1000 sqkm and publications per 100'000 capita. These estimates suggest that doubling the production value of cash crops is associated with 0.57-1.34 pp lower probability of exogamous marriages which amounts to 7.74-22.75% of the sample mean of the respective exogamy dummy. Until level 8 of the Ethnologue tree, absolute and relative effect sizes for the logged publications predictor are very close to the cash crop results just discussed. These substantive effect sizes are similar to or larger than logged distance to coast and between 0.45 and 1.32 times as large as the coefficient on a pre-colonial statehood dummy based on Murdock (1959, 1967). Both coastal distance and precolonial centralization are negatively and significantly associated with inter-ethnic marriage across most specifications.

Figure A9: Geographical persistence - cash crops and controls assigned to polygons & alternative publication data



Notes: The figure reports standardized beta coefficients from 16 OLS models with country-survey-round fixed effects. Standard errors are clustered at the survey location level. Each triangle represents the coefficient of the main variables of interest ((i) cash crop production per sqkm within the WLMS polygon the survey location is situated in and (ii) publications per capita in 1923 in the language of the WLMS polygon the survey location is situated in.)

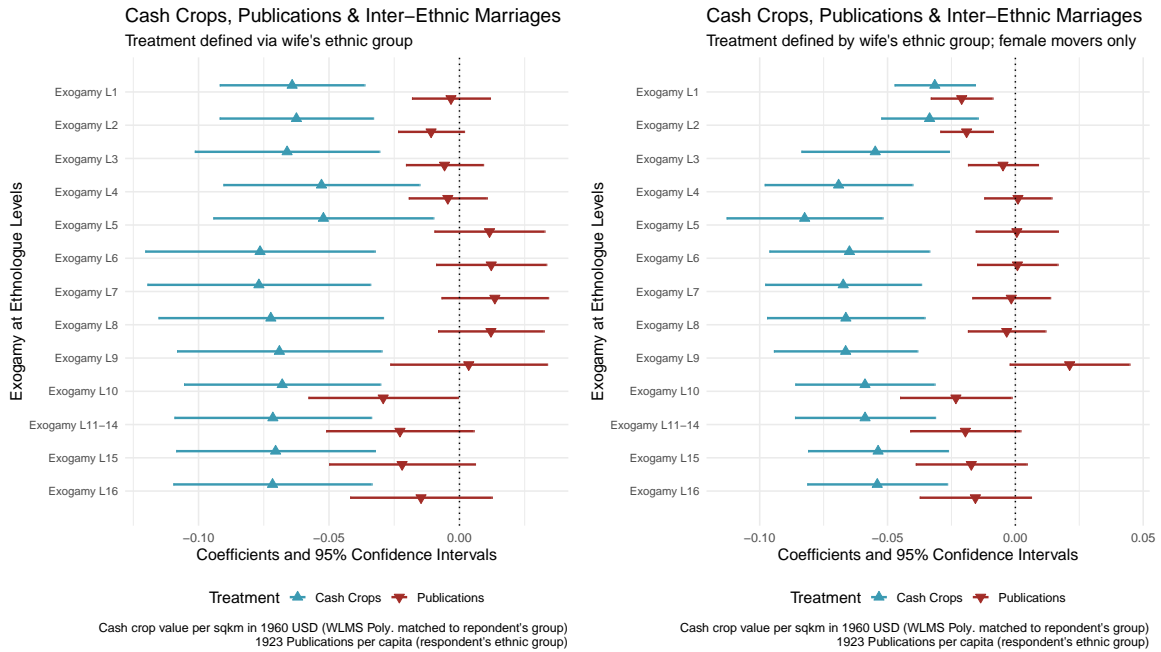
Figure A10: Geographical Persistence - Movers vs. Stayers Subsamples



Notes: The figure reports standardized beta coefficients from 16 OLS models with country-survey-round fixed effects. The dependent variables are binary indicators of inter-ethnic marriages at all levels of the Ethnologue language tree. The analysis sample is restricted to male stayers (top-left), female stayers (top-right), male movers (bottom-left) and female movers (bottom-right), respectively. Standard errors are clustered at the survey location level. Each triangle represents the coefficient of the standardized main variables of interest (i) cash crop production per sqkm within a radius of 15 km of each survey location is situated in and (ii) publications per capita in 1923 in the language of the WLMS polygon the survey location is situated in.)

Additional DHS results: Cultural Persistence

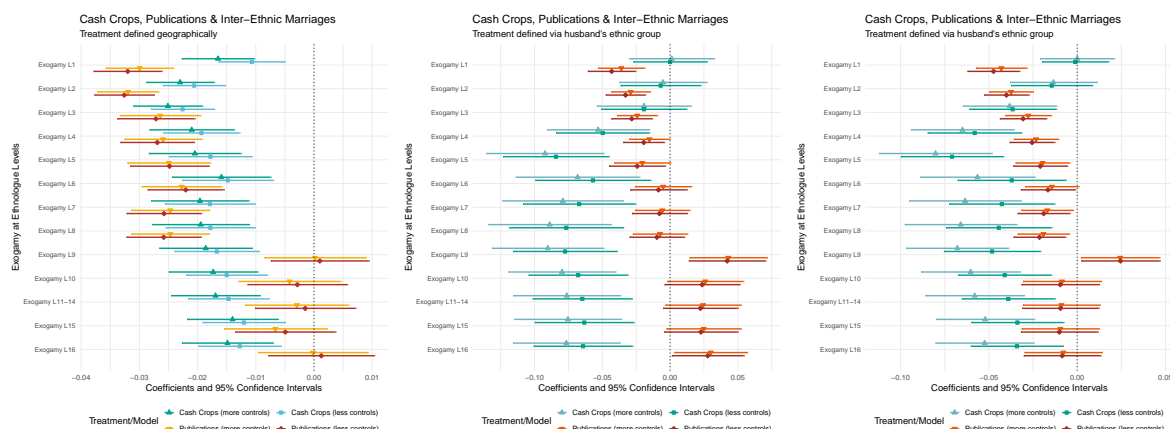
Figure A11: Cultural Persistence - Treatment assigned via wives' ethnicity



Notes: The figures reports OLS estimates from 16 models with survey location fixed effects. The dependent variables are binary indicators of inter-ethnic marriages at all levels of the Ethnologue language tree. Standard errors are clustered at the survey location level. Each triangle represents the coefficient of the main variables of interest: (i) the standardized USD value in cash crop production per sqkm within the WLMS polygon(s) matched to the wife's self-reported ethnic group (ii) standardized publications per capita in 1923 in the African language matched to the wife's self-reported ethnic group. Bars represent 95% confidence intervals. The left panel is based on analyses of the whole sample while the right panel reports results from models run on the subsample of ethnic movers only (i.e. wives who reside outside of the ethnic polygon of their self-reported ethnic group.)

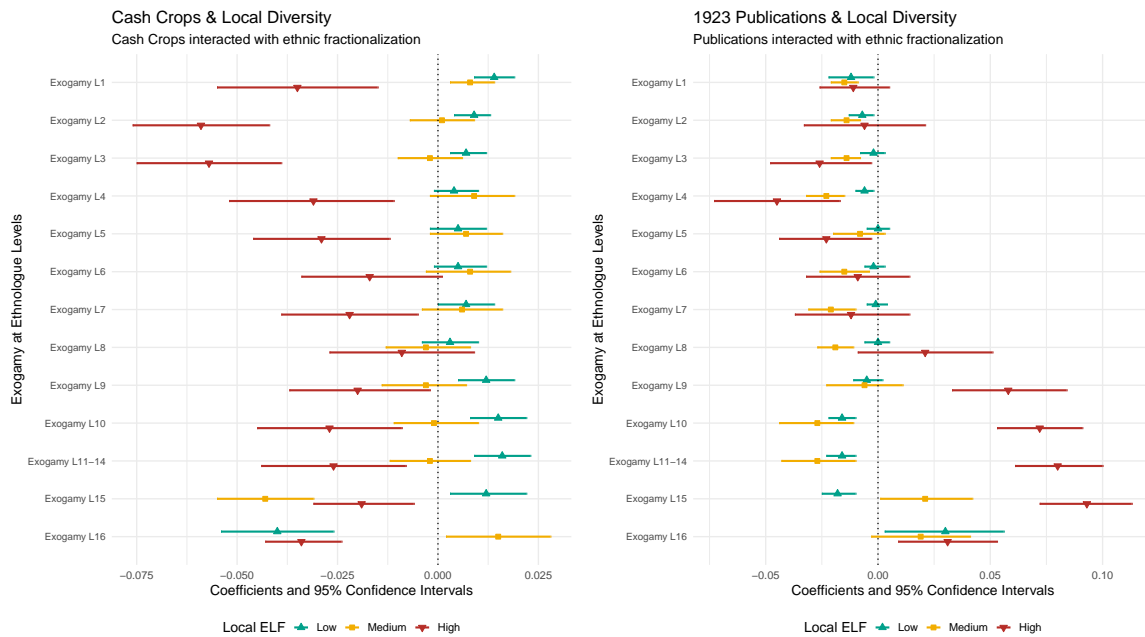
Additional DHS results: Robustness

Figure A12: Dropping Post-Treatment Controls



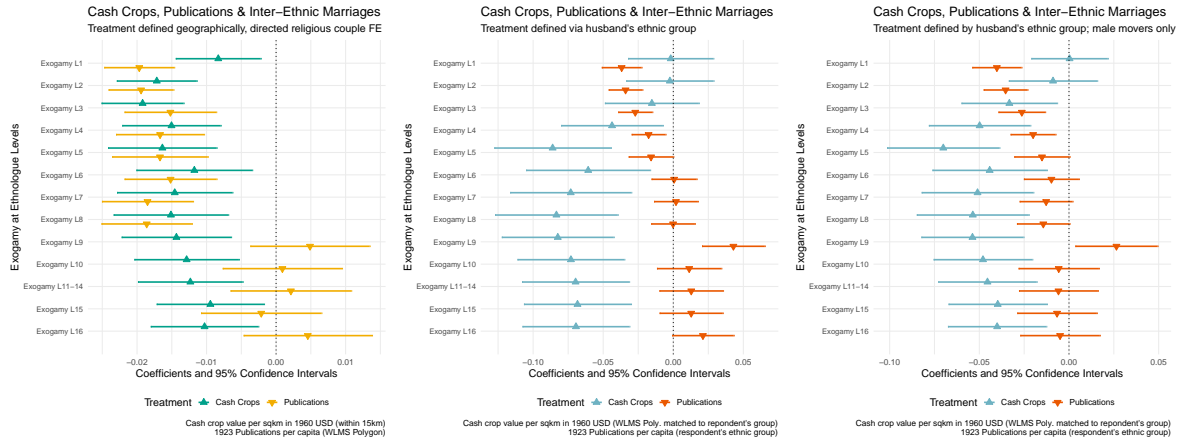
Notes: The figure replicates the geographic and cultural persistence analyses of inter-ethnic marriages from the main text (Figures 3 and 4). All three panels compare our baseline estimates to coefficients from models that drop all potential ‘bad controls’, i.e. male and female education level dummies, an urban dummy, and household wealth. Results remain very close to our baseline models. The cash crop coefficients get marginally smaller suggesting, perhaps, an alternative channel through which higher levels of development spurred by export agriculture minimally weakened the effects on ethnic boundaries.

Figure A13: Geographical Persistence: Local Ethnic Diversity?



Notes: The figure replicates the geographic persistence analysis of inter-ethnic marriages from the main text. We now interact the cash crop (left-hand panel) and publishing treatment (right-hand panel) with ethnolinguistic fractionalization scores calculated at the DHS enumeration area and at the same Ethnologue level of linguistic differentiation as the respective dependent variable. Instead of conventional linear interaction models, we use the recently proposed binning estimator (Hainmueller, Mummolo, and Xu 2019) to evaluate marginal effects of our treatments at typically high, intermediate, and low values of local ethnic fractionalization (i.e. at the median within each sample tercile of the moderating variable). Results indicate that, if anything, our findings are driven by relatively diverse survey locations. This further strengthens our confidence that ethnic competition between sons-of-the-soil and (historical) in-migrants rather than local-level ethnic homogeneity explains lower exogamy in historical cash crop areas.

Figure A14: Publications - Ethnic or Religious Boundary Making?



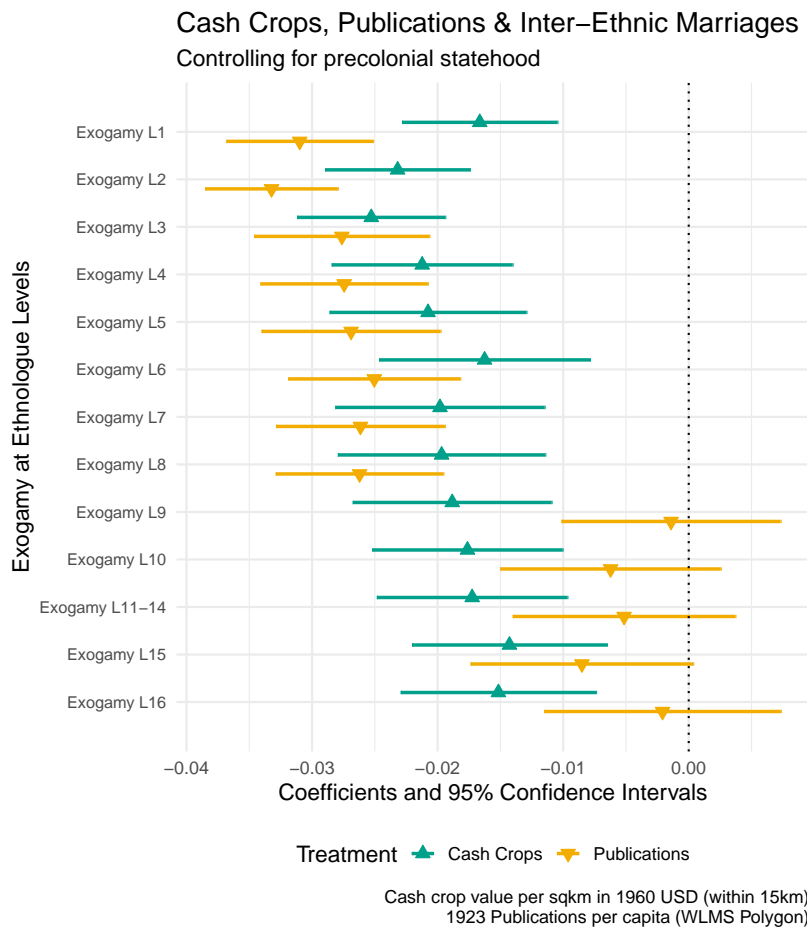
Notes: The figure replicates the geographic and cultural persistence analyses of inter-ethnic marriages from the main text. We add approx. 1000 directed religious couple fixed effects to test whether the ethnic marriage effects may plausibly driven by sharper religious boundaries. Directed fixed effects ensure that e.g. a Presbyterian women married to a Pentecostal men is not compared to a Pentecostal women married to a Presbyterian man. Results in the geographic persistence (left), cultural persistence (middle), and ethnic movers only models (right) remain very similar to our baseline specifications.

Figure A15: British vs. Non-British Colonies



Notes: The figure replicates the geographic and cultural persistence (leavers only) analyses of inter-ethnic marriages from the main text. Our historical treatments are now interacted with a binary indicator for former British colonies.

Figure A16: Controlling for Precolonial Statehood



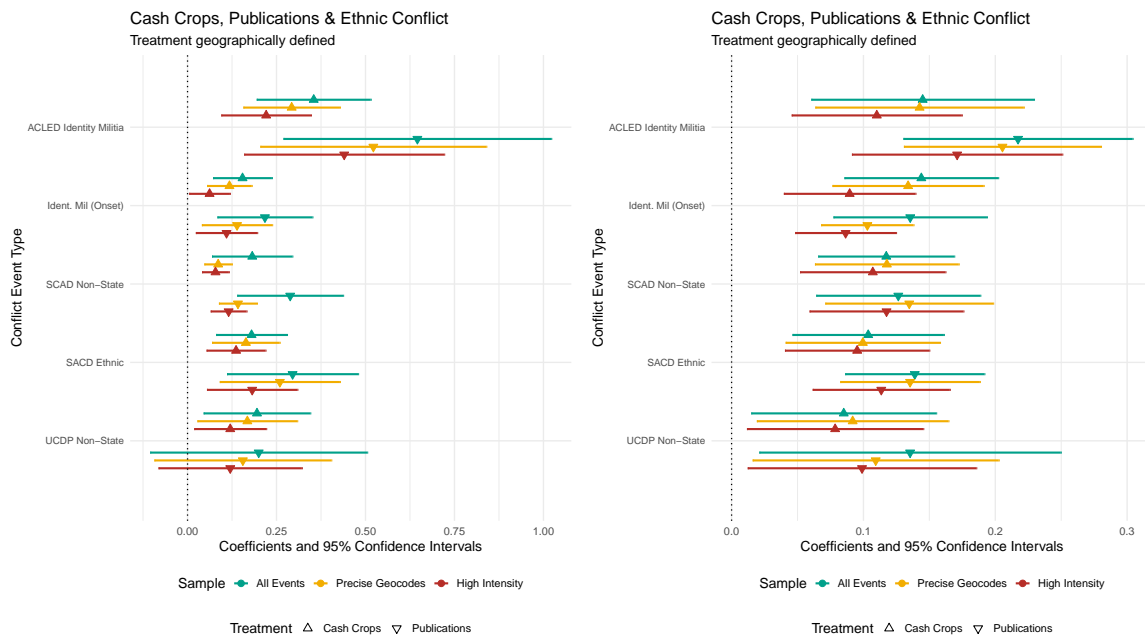
Notes: The figure replicates the geographic and cultural persistence analyses of inter-ethnic marriages from the main text. We add a precolonial statehood dummy based on Murdock (1967).

Ethnic Conflict

In a final set of analyses, we check whether the cash crop and printing and writing revolutions have left a legacy of local inter-group conflict. We spatially match plausibly ethnicity-related events from three prominent conflict data sets to the WLMS polygons used in the group-level analyses above. First, we count all ACLED events within WLMS polygons where one of the actors involved is designated as an “identity militia” (Raleigh et al. 2010). We also code an onset version that only counts the first event of the respective militia. Second, we use two conflict outcomes from the SCAD database (Salehyan et al. 2012): All events that SCAD codes as motivated by “ethnic discrimination, ethnic issues” and, alternatively, all events that are classified as “extra-governmental violence”, which captures local-level communal conflict. Third, we aggregate all communal/non-state violent events from the UCDP-GED database to WLMS polygons. We then rerun our baseline group-level OLS specifications with binary conflict indicators or logged event counts as dependent variables. To reduce concerns that any findings may be due to geocoding errors and/or reporting bias, we also run models in which we only aggregate precisely geocoded and/or high-intensity events to ethnic polygons (Weidmann 2016).

The coefficients on our historical cash crop and missionary publishing dummies are positive across all 30 specifications and statistically significant in all but three cases (publication dummies in the three models with the logged count of UCDP non-state events as outcome). These findings suggest that despite their differential impact on ethnic boundary-making as measured by inter-ethnic marriages, historical cash crop agriculture and printing similarly affect the current-day incidence of inter-group conflict.

Figure A17: Cash Crops, Publications & Contemporary (Ethnic) Conflict



Notes: The figure presents results from models of geocoded conflict events aggregated to WLMS polygons. The left-hand side uses logged event counts (+1) as dependent variables, whereas the right-hand side uses binary outcomes of whether any event is listed in polygon e since 1997 (ACLED) or 1989 (SCAD/UCDP). We use binary indicators for exposure to historical cash crop production and missionary publishing. All models include country fixed effects and confidence intervals are based on country-clustered standard errors. Control variables remain the same as in the group-level analysis above. We compare results across three subsets of conflict events to address concerns about geocoding errors and reporting bias: (a) All events, (b) only events with precise geocodes, and (c) only high-intensity events with precise geocodes (fatal events in ACLED/SCAD, > 5 fatalities in UCDP-GED.)

References

- Cagé, Julia, and Valeria Rueda. 2016. "The Long-Term Effects of the Printing Press in Sub-Saharan Africa." *American Economic Journal: Applied Economics* 8, no. 3 (): 1–31. (Cit. on p. [A4](#)).
- Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice." *Political Analysis* 27 (2): 163–192. (Cit. on p. [A24](#)).
- Hance, William A, Vincent Kotschar, and Richard J Peterec. 1961. "Source Areas of Export Production in Tropical Africa." *Geographical Review* 51 (4): 487–499. (Cit. on p. [A4](#)).
- Mann, Michael, and Valerie Sanders. 1994. *A bibliography of African language texts*. London: Hans Zell Publishers. (Cit. on pp. [A4](#), [A7](#), [A8](#), [A14](#)).
- Murdock, George Peter. 1959. *Africa: Its Peoples and Their Culture History*. New York: McGraw-Hill. (Cit. on p. [A19](#)).
- . 1967. *Ethnographic Atlas: a Summary*. Pittsburg University Press. (Cit. on pp. [A19](#), [A26](#)).
- Nunn, Nathan, and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review* 101, no. 7 (): 3221–3252. (Cit. on p. [A5](#)).
- Raleigh, Clionadh, et al. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47 (5): 651–660. (Cit. on p. [A27](#)).
- Rowling, Frank, and Charles-Edward Wilson. 1923. *Bibliography of African Christian literature*. London: Conference of Missionary Societies of Great Britain / Ireland. (Cit. on pp. [A4](#), [A13](#), [A14](#)).
- Salehyan, Idean, et al. 2012. "Social conflict in Africa: A new database." *International Interactions* 38 (4): 503–511. (Cit. on pp. [A3](#), [A27](#)).
- Weidmann, Nils B. 2016. "A closer look at reporting bias in conflict event data." *American Journal of Political Science* 60 (1): 206–218. (Cit. on p. [A27](#)).