Male Dominance and Cultural Extinction*

Eleonora Guarnieri Ana Tur-Prats

March 21, 2025

Preliminary and incomplete, please do not circulate or cite

Abstract: Why do some cultures and their associated values go extinct while others prevail? In this paper, we uncover a relationship between a society's deep-rooted gender norms and its risk of cultural extinction, proxied by language loss: languages from more gender-equal societies face a higher likelihood of extinction compared to those from male-dominant societies. We measure language status and male-dominance using the Ethnologue and the Male Dominance Index (Guarnieri and Tur-Prats, 2023), respectively, for a global sample of 4,750 languages. The negative relationship between male dominance and extinction holds after accounting for fundamental determinants of economic development and societal collapse at the language-group level such as geography, climate variability, conflict exposure, and historical factors, as well as after the inclusion of country fixed effects. We then leverage European colonization as a natural experiment to investigate how inter-group dynamics shape cultural extinction. In a dyadic framework, we find that Indigenous societies with more gender-equal norms than their colonizers are significantly more vulnerable to cultural extinction. Cultural distance in gender norms is a stronger predictor of extinction than linguistic distance, distance in pre-colonial institutions, or the characteristics of either the colonizer or the Indigenous group.

Keywords: Cultural diversity, Language extinction, Gender norms, Colonialism, Inter-group contact

JEL classification: Z10, O10, F54, J16

*Author affiliations and contacts: Guarnieri (University of Bristol and CESifo, eleonora.guarnieri@bristol.ac.uk), Tur-Prats (University of California at Merced and IZA, atur-prats@ucmerced.edu). We thank numerous seminar and conference participants at Warwick, Queen Mary, ECARES, Luxembourg, St Gallen, Geneva, Ca' Foscari University of Venice, Bristol, Birmingham, Rotterdam, UC Merced, Applied Micro Workshop at Padua, and the CEPR Annual Symposium for useful comments and suggestions. Tiffany Ho and Johanna Lozano-Lavao provided excellent research assistance. All errors remain our own.

1 Introduction

Why do some cultures and their associated values go extinct while others prevail? Unraveling this question is key to understanding the global patterns of cultural diversity and the forces that shape them. Cultural extinction can result from the physical loss of individuals, the collapse of a social entity, or the abandonment of culture-specific practices (Boyd and Richerson, 1988; Zhang and Mace, 2021). These processes, in turn, may be triggered by internal dynamics within a group, inter-group competition, or both (Zhang and Mace, 2021). In this paper, we show that a society's risk of *cultural extinction* is shaped by its deeply rooted gender norms.

Studying cultural extinction is inherently challenging, as determining whether and when a culture has disappeared is complex. We address this challenge by employing language extinction as a proxy for cultural extinction. Language is an important carrier of culture, transmitted intergenerationally within populations, and oftentimes used to define cultural and ethnic boundaries. The loss of a language represents the disappearance of a fundamental cultural asset. In the late modern era, cultural extinctions have reached an unprecedented scale: nearly half of the world's approximately 7,000 languages are either endangered or already extinct (Eberhard *et al.*, 2024), with projections suggesting that up to 50% could vanish in the next century (Crystal, 2002). Like cultural extinction, language loss occurs when the speaking population disappears (as seen with certain Indigenous groups in the Americas), undergoes voluntary or forced cultural assimilation (such as Cornish speakers shifting to English), or faces a combination of these pressures (Nettle and Romaine, 2000).

Gender norms are not only a prominent part of culture, but also reflect key aspects of a society's economic and institutional structure. Our measure of deep-rooted gender norms, the male dominance index (MDI) (Guarnieri and Tur-Prats, 2023), is based on ten ancestral ethnic characteristics from the Murdock Ethnographic Atlas (EA) related to subsistence activities and family arrangements that have shaped the role of men and women in society and the economy. Our hypothesis is motivated by the striking empirical pattern shown in Figure 1: language extinction is most prevalent among the world's most gender-equal groups, suggesting that gender norms may play a crucial yet overlooked role in the process of cultural evolution.

Why might gender norms influence a society's chances of survival? First, they may

correlate with known deep-rooted geographical factors that either promote or hinder development—for example, gender-equal groups are concentrated closer to the Equator, which in turn correlates with economic development. Second, gender norms may co-evolve with societal traits that affect long-term survival, such as fertility rates. Third, gender norms can shape inter-group dynamics, influencing a group's expansionist tendencies or its resilience against external threats—male-dominant groups, for instance, may be more likely to develop stronger military forces. Building on this framework, we empirically examine whether and how gender norms impact the cultural extinction of human societies by considering two key dimensions: the *absolute* level of male dominance within a cultural group and its *relative* male dominance compared to competing cultures.

To this end, we assemble a novel dataset that combines information on languages and their status from the Ethnologue (Lewis, 2014) with the gender norms of the societies speaking those languages, captured by the male dominance index (MDI) (Guarnieri and Tur-Prats, 2023). To assign a level of male dominance to each linguistic group, we follow the methodology of Bahrami-Rad *et al.* (2021), which links languages to ethnic groups in the Murdock EA based on their proximity in the linguistic tree. With this data at hand, we start by regressing an indicator of whether or not a language is extinct on the MDI. We find that, the more male dominant a society, the less likely is its language to have gone extinct: a one standard deviation increase in male dominance is associated with a 8.6 percentage points decrease in the likelihood of cultural extinction.

Cultural norms about gender are likely shaped by deep determinants of comparative development and may co-evolve alongside other historical and institutional arrangements that shape a society's chances of survival. To examine whether gender norms influence extinction independently of these factors, we enrich our data with a large set of geo-climatic characteristics at the language level such as absolute latitude, elevation, ruggedness, distance to the coast, caloric suitability, disease environment and natural disasters, and historical determinants of economic development such as historical conflict, settlement patterns and political complexity. The relationship between gender equality and extinction remains remarkably strong and significant after sequentially including these controls. Our results are also robust to the inclusion of country fixed effects—which nets out the effect of national history and institutions—, alternative measures of male dominance and extinction, the inclusion of additional controls, and a restricted sample limited to Indigenous languages.

In a next step, we acknowledge that a culture's survival is not determined in a vacuum. In an interconnected world, cultural groups interact and compete against each other; as a result of this contact, some cultures may perish while others may thrive. To study how inter-group relationships influence cultural extinction in modern history, we focus on European colonization as a historical natural experiment. This large-scale shock exposed most non-European groups in the world to a new competing dominant culture. Yet, there remains variation in the status of Indigenous cultures subjected to a given colonizer: some cultures survived, while others did not. This provides us with an ideal setting to test whether a group's cultural norms *relative* to those of a competing culture influence its likelihood of extinction.

In a dyadic framework, we relate each Indigenous language to its colonizer and find that a greater absolute distance in gender norms between the colonizer and the Indigenous group increases the latter's likelihood of extinction. Importantly, we find that cultural distance in gender norms from the colonizer—i.e., a group's degree of male dominance *relative* to a competing culture—is a stronger predictor of an Indigenous group's extinction than the *absolute* level of male dominance of either the Indigenous group itself or the colonizer, underscoring the importance of inter-group contact as a driver of extinction. When exploring this association further, we find that extinction is driven by a specific clash, one where the Indigenous language is more gender equal than the colonizer, a finding that is robust to the inclusion of colonizer fixed effects.

We then delve into the interpretation of our findings. Because gender norms are a key component of culture, we want to ascertain whether cultural extinction is a function of a mismatch in gender norms or of a more general cultural mismatch. We thus compute the linguistic distance—a broadly-used measure of cultural distance— between the colonizer and the Indigenous groups, and check whether it predicts cultural extinction. We do not find any significant effect. However, when including both the linguistic distance and the distance in gender norms, the effect of the latter maintains its size and statistical significance.

Differences in political complexity might influence extinction if more hierarchical and centralized groups are able to conquer less politically-complex groups through greater military advantage (Zhang and Mace, 2021). We test whether differences in (pre-colonial) political complexity between the colonizer and the Indigenous groups are a predictor of cultural extinction. We find a relatively small but statistically significant effect of distance

in pre-colonial political complexity, measured from the Ethnographic Atlas. However, when including both distance in political complexity and in gender norms, the coefficient of distance in political complexity decreases substantially both in magnitude and statistical significance while the coefficient of distance in gender norms barely changes.

These results point us to a gender-based explanation for cultural extinction. To better understand why a more pronounced cultural mismatch in gender norms between the colonizer and the Indigenous group results in a higher likelihood of cultural extinction, we rely on historical narratives and theories from other social sciences. Philosopher Maria Lugones places gender in a central position to understand the consequences of colonialism, claiming that "understanding the place of gender in pre-colonial societies is pivotal to understanding the nature and scope of changes in the social structure that the process constituting colonial/modern Eurocentered capitalism imposed [...]" (Lugones, 2016, p.16). Focusing on the impact of colonization on Native American groups, Paula Gunn Allen describes how European missionaries and colonizers could not tolerate that women occupy prominent positions, and that "the physical and cultural genocide of American Indian tribes is and was mostly about patriarchal fear of gynocracy" (Allen, 1992, p.19).

European colonizers not only brought their institutions and technology, but also imported their male-dominant social and economic structure into the territories they colonized. As described by Perdue (1998) for the Cherokee, European colonizers established that men had to replace women as farmers, undermined the principle of matrilineal kinship, outlawed polygamy, and imposed the nuclear family structure, among other policies aimed at "civilizing" the Cherokee people. These disruptions are likely to have contributed to the breakdown of the most gender-equal Indigenous groups' social fabric and the collapse of their economies, ultimately leading to their cultural extinction.

This paper contributes to various strands of literature. By proposing and empirically testing a novel hypothesis on the determinants of language extinction, we add to the interdisciplinary scholarship on language loss (Nettle and Romaine, 2000; Crystal, 2002; Austin and Sallabank, 2011; Fishman, 1991; Fernihough *et al.*, 2024; Blouin and Dyer, 2022). Specifically, we identify deep-rooted gender norms as a key driver of language extinction and demonstrate that cultural clashes over gender norms with a dominant group place more gender-equal language groups at heightened risk of linguistic extinction. Furthermore, our systematic, global analysis of language extinction expands the scant empirical analyses on this phenomenon (Bromham *et al.*, 2022; Amano *et al.*, 2014; Kik

et al., 2021).

More broadly, this paper contributes to the literature on ethnolinguistic diversity and the evolution of human cultures. While previous research in economics has explored the formation of cultural groups (Ahlerup and Olsson, 2012; Michalopoulos, 2012), the evolution of cultural values within existing groups (Giuliano and Nunn, 2021; Giavazzi *et al.*, 2019; Bisin and Verdier, 2000; Abramitzky *et al.*, 2020; Fouka, 2020), and the determinants of intra-group cultural heterogeneity (Galor *et al.*, 2024), our study is the first to focus on the *disappearance* of established cultural entities. By reintroducing now-extinct groups into the study of culture and explicitly acknowledging the phenomenon of cultural extinction, we provide new insights into a critical factor shaping contemporary levels of ethnolinguistic diversity and the intergenerational transmission of cultural traits.

Finally, our paper adds to the literature on the consequences of European colonization. While the economic and institutional effects of colonization have been extensively examined in the economics literature (Acemoglu *et al.*, 2001), its influence on the world's cultural landscape remains less understood. We demonstrate that the cultural survival of Indigenous groups was shaped by the extent to which their gender norms diverged from those of the colonizers. In doing so, we provide the first empirical evidence highlighting the central role of gender relations in determining the cultural outcomes of European colonization.

2 Conceptual Framework and Hypotheses

Linguists have identified multiple causes of language endangerment and eventual extinction (Nettle and Romaine, 2000; Crystal, 2002; Austin and Sallabank, 2011). These causes can be broadly categorized into *internal factors*—arising within the language-speaking community—and *external factors*—resulting from contact with other linguistic groups. Often, these factors interact in complex ways, making it difficult to isolate a single cause. At the language-group level, natural catastrophes, famine, and disease, can severely reduce populations, displace communities, and contribute to cultural collapse.[1] More commonly, language endangerment arises from contact with other linguistic and cultural groups. These inter-group dynamics often create a power imbalance between a

[1] Austin and Sallabank (2011) mention the case of the Malol in Papua New Guinea and the populations inhabiting the Andaman Islands, devastated by massive tsunamis in 1998 and 2004, respectively. The resulting destruction, loss of life, and population displacement contributed to the extinction of their language.

dominant and a threatened cultural group. The cultural, political, and economic dominance of the more powerful group can prompt a language shift within the threatened group through various channels. These may include national policies, such as language bans or lack of political representation. More extreme drivers of language or cultural extinction involve war, genocide, repression, or forced displacement.[2]

We argue that gender norms can play a role in both the internal and external factors that contribute to cultural extinction. Specifically, we hypothesize that: (i) societies with more gender-equal economic structures are more susceptible to cultural extinction, and (ii) the greater a society's gender equality relative to a competing culture, the higher its likelihood of cultural extinction.

Gender norms can matter for the *internal dynamics* that shape a group's survival in two ways. First, gender norms might correlate with, or be the result of, structural factors that influence economic development and thus a society's survival prospects. Gender norms may be shaped by ecological and geo-climatic conditions, as well as exposure to natural disasters and environmental shocks, all of which can directly either promote or hinder development. For instance, gender-equal societies are concentrated closer to the Equator; at the same time, distance from the Equator correlates with economic performance (Gallup *et al.*, 1999; Hall and Jones, 1999). Second, gender norms may co-evolve with or favor societal traits that affect long-term survival. Specifically, male-dominant societies may develop certain socio-demographic behaviors (e.g., fertility patterns), institutions (e.g., larger armies), or other cultural traits that reinforce the intergenerational transmission of language, reducing the likelihood of cultural extinction.

Gender norms can also shape extinction via *inter-group dynamics*. Gender-equal societies might be more susceptible to the negative consequences of disruptive inter-group relationships, including colonization, economic and cultural exclusion, or war. In these contexts, gender-equal groups may be more prone to assimilation, subjugation, or eventual cultural extinction. For instance, research demonstrates that male-dominant groups are more likely to use sexual violence in conflict, especially against more gender-equal opponents (Guarnieri and Tur-Prats, 2023).

We take these hypotheses to the data and test whether (i) the *absolute* level of male dominance of a cultural group is associated with its likelihood of extinction net of known

[2]For instance, the colonial genocide in Tasmania led to the almost total annihilation of their Indigenous population.

geographical, ecological, and institutional factors that are important for comparative development; and (ii) extinction is shaped by a groups' *relative* male dominance compared to a competing culture. To test this second hypothesis, we leverage the European colonization, a natural experiment of modern history that exposed the majority of non-European groups in the world to a new dominant and competing culture.

3 Data

To study the relationship between gender norms and cultural extinction, we build a novel dataset at the language-country level combining information on language vitality, gender norms, geo-climatic characteristics and historical arrangements.

3.1 Languages and their Status

We begin by obtaining a list of all known human languages from the ISO 639-3 database curated by SIL International. Every "distinct human language that has been documented, whether living, extinct, or constructed, and whether its modality is spoken, written or signed" (SIL International, 2023) is uniquely identified by a three-letter code, with a total of 8,552 entries in the database. ISO 639-3 defines language boundaries based not only on linguistic criteria (e.g., intelligibility), but also considerations pertaining to shared versus distinct ethno-linguistic identity among speakers, an appealing feature for the specific purpose of this paper.[3] Only so-called "full languages" are eligible to receiving a code, i.e., those that are used in a variety of domains and support communication across all genders and all ages.[4]

To retrieve information on the status or vitality of languages and construct a measure of cultural extinction, we exploit the Ethnologue database (Eberhard *et al.*, 2024), one of the most authoritative sources of information on languages. Building on the ISO 639-3 classification, the Ethnologue assigns to each language spoken in a given country one out of 13 possible values of the so-called Expanded Graded Intergenerational Disruption Scale (EGIDS) (Lewis and Simons, 2017).[5] The higher the value in the EGIDS scale, the

[3]In general, ISO 639-3 does not categorize dialects—defined as divergent varieties of a language with which they share similar lexical, phonological or grammatical features—as separate entities.

[4]Based on this definition, avoidance languages, secret languages, ceremonial languages, play languages, urban languages, youth languages, special-purpose languages are not within the scope of this classification (SIL International, 2023).

[5]Because it does not cover ancient, constructed, and historical languages, the Ethnologue comprises 7,678 of the 8,552 languages listed in the ISO 639-3 database.

greater the disruption to the intergenerational transmission of the language.

Table A-1 illustrates the various categories a given language in a specific country can belong to. Lower values are assigned to national languages (e.g., English in the United States, Spanish in Mexico, Khmer in Cambodia, Swahili in Kenya) or languages that are in vigorous use (e.g., Basque in Spain, Krio in Sierra Leone, Hindi in Nepal, Italian in Argentina). Conversely, higher values indicate increasing degrees of endangerment. We code a language in a given country as *extinct* if it belongs to one of the following categories: *nearly extinct*, i.e, the only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language (e.g., Dakota in Canada); *dormant*, i.e., languages that serve as a reminder of heritage identity for an ethnic community but lack proficient speakers (e.g., Karuwali in Australia); and fully *extinct*, i.e., languages that are no longer used and lack speakers retaining any ethnic identity associated with them (e.g., Puquina in Peru).

3.2 Gender Norms

To assign a measure of male dominance to each language group, we match languages to the male dominance index (MDI) developed by Guarnieri and Tur-Prats (2023). This index is based on ancestral ethnic characteristics from the Murdock Ethnographic Atlas, which describes cultural practices for more than 1,200 societies measured prior to industrialization and European contact. Informed by interdisciplinary literature on the origins of societal gender norms, Guarnieri and Tur-Prats (2023) select ten key ethnic traits related to family arrangements, subsistence activities, and inheritance patterns that have been associated with notions of gender equality or male dominance. The more central the role of women in these domains in ancestral times, the more gender-equal the cultural norms of a society.

The selected ethnic traits—matrilineality, polygyny, dependence on fishing, gathering, non-herding animal husbandry, shifting agriculture, bride price, plough use, nuclear family structure, and dependence on pastoralism—are combined into a single index using principal component analysis, with the first principal component serving as the male dominance index (MDI). This component loads positively on traits traditionally associated with male dominance and negatively on those linked to greater gender equality. Guarnieri and Tur-Prats (2023) show that this index tallies well with survey-based, contemporary measures of gender equality and inequality including female employment, son preferences, gender

attitudes, and domestic violence.

To assign the MDI to languages, we extend the method proposed by Bahrami-Rad *et al.* (2021) to extinct languages. First, we use the D-PLACE database, which maps 1,209 languages to 1,291 societies in the EA, to create what we refer to as "direct links" between languages and EA groups. Languages directly linked to an EA group receive the respective group's MDI. To languages not directly linked to the EA, we assign the MDI of their nearest direct link(s) in the linguistic tree. Appendix B.1 describes this procedure in detail.

3.3 Geo-climatic and Historical Covariates

To isolate the role of gender norms in explaining extinction, we enrich our dataset with a large set of language-level geo-climatic and historical covariates. Additional details on data sources and matching are described in the Appendix.

Geography. Information on the locations where languages are spoken comes from the Ethnologue map, which provides polygons for 7,626 living language-countries (or 6,425 distinct languages). For the remaining languages not featured in the map, including extinct ones, we retrieve coordinates from the Glottolog (Hammarström *et al.*, 2023) and other sources. We exploit this information to compute the average elevation and ruggedness across each language polygon (or a 0.25 degrees radius around the point location for cases in which information on the language's polygon is unavailable) using the GTOPO30 rasters (U.S. Geological Survey, 1996); the average caloric suitability using Galor and Özak's (2016) index; and distance to the coast using Natural Earth's (2010) coastline shapefiles.

Climate, Disease Environment, Natural Resources, and Disasters. The next set of controls comprises various geo-climatic indicators. First, we build a measure of climate variability at the language level based on Mann *et al.*'s (2009) data on the average annual temperature anomalies from 500 to 1900, which ranges from -1.54 to 1.1. Second, we compute the ratio of a language's area located in tropical climate using the Köppen-Geiger map (Beck *et al.*, 2018). Third, to capture a language's disease environment, we take the average of Kiszewski *et al.*'s (2004) malaria index (which ranges between 0 and 39) across a language's area. Fourth, we measure the presence of natural resources using the PRIO-grid dataset and building a binary indicator equal to 1 if gold, petroleum, diamonds, or gems resources are present in a language polygon/radius and 0 otherwise. Fifth, we

use the EM-DAT database (Centre for Research on the Epidemiology of Disasters, 2023), which records the occurrence of natural disasters worldwide since 1900, to quantify the exposure of each language group to such events.

Historical Controls. Finally, the last set of controls comprises historical conflict and other societal characteristics. Our measure of historical conflict relies on Miller's (2022) Historical Conflict Event Dataset and captures the number of battles a linguistic group was exposed to from 1468 BC to the 18th century. Information on settlement patterns (nomadic vs sedentary) and a society's pre-colonial political complexity (ranging 0-4, from no political authority beyond the community to large states) comes from the EA, as does the MDI.

3.4 Summary Statistics

The unit of observation in the resulting dataset is a language-country. Our estimation sample comprises 4,750 languages spoken in 172 countries (N=5,763). Table A-2 in the Appendix reports summary statistics. 11 percent of the languages-country in our sample are extinct according to our definition. Among these 596 languages (corresponding to 617 observations), 29 are "dormant" according to the EGIDS scale, 34 percent are "nearly extinct" and 37 percent are fully "extinct."

For a subset of dormant and extinct languages, the Ethnologue provides some information on the timing of extinction, approximated by the death of the last known speaker. Figure A-1 and A-2 in the Appendix show the distribution of extinctions over time, for all the languages in the Ethnologue and for those in our estimation sample, respectively.[6] As can be seen, most languages have disappeared over the last two centuries.

4 Gender Norms and Cultural Extinction

4.1 Empirical Strategy

To estimate the relationship between a society's gender norms and its likelihood of extinction, we estimate the following linear probability model:

$$\text{extinct}_{ic} = \lambda \text{MDI}_i + \gamma' \mathbf{X}_{ic} + \epsilon_{ic} \tag{1}$$

[6]Information on the timing of extinction is available for 383 out of 414 dormant or extinct languages in our estimation sample.

where extinct_{ic} is an indicator that equals 1 if a language i in country c is either nearly extinct, dormant, or extinct, and 0 otherwise; MDI_i is the male dominance index, a continuous variable ranging between 0 and 1 capturing the degree of male dominance of language i.[7]

$\mathbf{X}_{ic}$ is a vector of language-by-country control variables, encompassing a comprehensive set of factors that could potentially confound the analysis. These controls are motivated by existing literature that links them to economic development, ethnolinguistic diversity, and societal collapse.

Geographical characteristics. Our first control is absolute latitude (or, equivalently, distance to the Equator). Based on Gallup *et al.*'s (1999) and Hall and Jones's (1999) studies documenting a correlation between distance to the Equator and economic performance, absolute latitude is commonly employed in the literature on comparative development to account for geographic effects (see, e.g., Acemoglu *et al.* (2001); Michalopoulos (2012); Rodrik *et al.* (2004)). We also account for distance to the coast and ruggedness to proxy for trade access and isolation, factors that may underlie economic development (Nunn and Puga, 2012; Gallup *et al.*, 1999) as well as language extinction directly. Controlling for elevation and the caloric suitability index allows us to account for confounders related to agricultural productivity, another fundamental determinant of economic development (Galor and Özak, 2016) and ethnic diversity (Michalopoulos, 2012).

Climate and disease. Nettle (1998) highlights that a key factor explaining the global distribution of linguistic diversity is climate variability, due to its effect on the seasonality of food production. Tropical climate and the burden of disease are at the roots of differential income levels and growth through their impact on productivity or institutions (Gallup *et al.*, 1999; Acemoglu *et al.*, 2001). We therefore add climate variability, tropical climate, and the malaria index to our set of controls.

Natural resources and natural disasters. We employ an indicator for the presence of natural resources (gold, petroleum, diamonds, or gems) and the log number of natural disasters as controls. The presence of natural resources has been shown to influence institutions, economic growth, and violence (Humphreys, 2005; Ploeg, 2011), factors that may matter for cultural extinction as well. Moreover, the occurrence of natural disasters is directly linked to a society's risk of collapse (Diamond, 2011; Tainter, 1988).

[7]Notice that the MDI is specific to a language, not a language-country. This is because the unit of observation in the Murdock Ethnographic Atlas, on which the MDI is based, is an ethnic group, irrespective of whether the group is located in multiple countries or not.

Exposure to conflict. To account for a language's exposure to historical conflict, a driver of long-run development (Besley and Reynal-Querol, 2014), we control for the log number of battles in a language's polygon.

Precolonial arrangements and institutions. Finally, to disentangle the role of cultural norms in explaining extinction from that of institutions (Alesina and Giuliano, 2015; Bisin and Verdier, 2017), we control for the ancestral political complexity of a society and for its settlement patterns (i.e., sedentary vs nomadic), factors that may co-evolve with gender norms and define societies' development trajectories (see, e.g., the role of pre-colonial ethnic institutions in influencing contemporary comparative regional development in African countries described by Michalopoulos and Papaioannou (2013)).[8]

We include these sets of controls sequentially, from the arguably most exogenous and predetermined (e.g., geography) to the more endogenous ones, i.e., those that are likely co-evolving with gender norms (e.g., exposure to historical conflict and pre-colonial institutions). In the baseline specification, we cluster standard errors at the country level.

4.2 Results

Table 1 displays estimates of λ. We find a strong and significant relationship between gender norms and the likelihood of extinction, confirming the visual pattern observed in Figure 1: the more male dominant a society's deep-rooted cultural norms, the lower its likelihood of extinction. Column 1 reveals that moving from the most gender equal to the most gender-unequal society in our sample decreases the probability of extinction by 48 percentage points. Equivalently, a one standard deviation increase in male dominance is associated with a 0.28 standard deviations lower likelihood of extinction.

This sizable estimate may be capturing not only the effect of gender norms, but also of confounding factors that may correlate with gender norms and economic development. For instance, it could be that gender-egalitarian norms have systematically emerged from geographical features less conducive to long-run development, from the occurrence of geo-climatic shocks putting societies at higher risk of extinction, or from the absence of growth-inducing natural resources. If that was the case, our estimate would be picking up the effect of these confounders; as such, we would expect the size and significance of the MDI coefficient to be substantially affected by the inclusion of these factors as

[8]Like the MDI, these controls are taken from the Murdock Ethnographic Atlas and are therefore specific to a language (not a language-country).

controls. To the contrary, in columns 2-5 we find that the MDI remains a remarkably strong predictor of extinction after netting out the effect of potential geographical and geo-climatic confounders.

Similarly, gender norms may co-evolve together with institutional or organizational features of a society that may be crucial for its survival such as the degree of political complexity, settlement patterns, or exposure to war. While acknowledging the presence of a potential "bad controls" issue—given that these dimensions are arguably more endogenous—we explore whether the MDI coefficient survives to the inclusion of these controls. As can be seen in columns 6 and 7, the estimate remains large and statistically significant: in column 7, a one standard deviation increase in the MDI decreases the likelihood of cultural extinction by 0.2 standard deviations.

Remarkably, the relationship between gender norms and extinction also holds when comparing languages within the same country, as shown in Table 2. The effect is smaller in magnitude compared to the baseline specification, but remains sizable: when including the full set of controls, a one standard deviation increase in the MDI decreases the likelihood of extinction by 0.1 standard deviations. Including country fixed effects allows us to isolate the role of gender norms from that of national institutions, colonial history, or policies common to all languages in a country.

4.3 Robustness

Each of the single ethnic characteristics that form the MDI may affect cultural extinction directly and not (only) through gender norms. To assess whether the effect is driven by single ethnic traits rather than the MDI, we take two steps. First, in Table A-4, we run a horse-race between the MDI and each ethnic trait and find that the MDI remains a strong predictor of extinction. Second, we replicate our findings using an alternative measure of gender norms at the language level. Using folkloric texts, Michalopoulos and Xue (2021) develop a measure of male dominance bias (MDB), which captures the degree to which an ethnic-specific oral tradition describes men as more violent, less dependent, less engaged in domestic affairs, and more physically active relative to women. Michalopoulos and Xue (2021) show that this measure correlates with the historical and contemporary role of women. Table A-5 displays a strong correlation between the MDB and our MDI, which further validates our measure of gender norms.[9] In Table A-6, we find that our main

[9]Conveniently, Michalopoulos and Xue (2021) provide the MDB at the Murdock ethnic group level.

results hold when employing the MDB rather than the MDI: the more male dominant the oral tradition of a society, the less likely its cultural extinction.

Our results are robust to a large battery of additional checks. In Table A-7 we show that results hold when restricting the sample to languages that are indigenous to a country (columns 1 and 2), when employing an alternative methodology to match the MDI to languages[10] (columns 3 and 4), and when only focusing on languages that can be directly matched to a Murdock ethnicity (columns 5 and 6). Because our MDI stems from ethnic characteristics coded by an ethnographer potentially observing societies at different points in time—and this, in turn, may be endogenous to a society's survival or extinction—we show that our estimates remain significant and large when controlling for the year in which a given group is observed (columns 7 and 8).

Table A-8 shows that results also hold when employing wider or narrower measures of extinction as dependent variables, i.e., one that includes shifting and moribund languages (columns 1 and 2) and one that only considers fully extinct languages (columns 3 and 4), respectively. Columns 5 and 6 remove from the sample languages that have gone extinct before 1900 (and thus were less likely to be observed by ethnographers), while columns 7 and 8 show that results hold using a dataset collapsing the data to the language level.

Taken together, this set of results points to a robust and strong relationship between gender norms and cultural extinction, operating over and above the confounding effect of other societal-level factors that may matter for both the formation of its gender norms and its cultural survival.

5 Gender Norms Distance, Colonization, and Cultural Extinction

We now turn to exploring how cultural norms about gender influence cultural extinction in the context of inter-group dynamics. To this end, we exploit a natural experiment of modern history—European colonization—that exposed the majority of non-European cultural groups to a new competing dominant culture. Colonization is described by linguists as one of the major causes of cultural extinction in recent times (Mufwene, 2002) and, according to Lugones (2016), gender relations in pre-colonial societies are key for understanding the scope of the social transformation enacted by European colonization:

> Understanding the place of gender in pre-colonial societies is pivotal to under-

[10] Appendix B.1 describes this alternative MDI in detail.

> standing the nature and scope of changes in the social structure that the processes constituting colonial/modern Eurocentered capitalism imposed. Those changes were introduced through slow, discontinuous, and heterogenous processes that violently inferiorized colonized women. The gender system introduced was one thoroughly informed through the coloniality of power. Understanding the place of gender in pre-colonial societies is also pivotal in understanding the extent and importance of the gender system in disintegrating communal relations, egalitarian relations, ritual thinking, collective decision making, collective authority, and economies. —Maria Lugones (2016), *The Coloniality of Gender*, p. 16.

Among the Indigenous groups subjected to the same colonizer, there remains large variation in the extent to which languages have undergone extinction. For instance, while the Native American Keres language is still spoken in New Mexico by members of the ethnic community, the Native American Piro language—which used to be spoken in the very same region—is now extinct.

5.1 Empirical Strategy

To understand this variation, we restrict the sample to 3,931 Indigenous languages in 106 countries (N=4,346) affected by European colonization.[11] Table A-3 presents summary statistics for this sample. On average, the prevalence of extinction is higher compared to that in the full sample of languages described in Table A-2, while the level of male dominance is lower.

We relate each Indigenous language to their colonizer in a dyadic setting and consider the role of cultural distance in gender norms in explaining extinction. Data on the identity of the colonizer comes from the Colonial Dates Dataset (COLDAT) dataset (Becker, 2019). After assigning to each colonizer a measure of male dominance via language, we calculate the absolute difference in gender norms between each Indigenous language i and the colonizer of its respective country c as follows:[12]

$$\text{MDI}_{ic}^{Dist} = |\text{MDI}_i - \text{MDI}_c|$$

[11] The Ethnologue provides information on whether or not a language spoken in a country is indigenous.

[12] Whenever a linguistic group was colonized by multiple colonial powers, we take the average MDI of all the colonizers of a given country.

Next, we estimate the following specification:

$$\text{extinct}_{ic} = \theta \text{MDI}_i + \delta \text{MDI}_{ic}^{Dist} + \rho' \mathbf{X}_{ic} + \epsilon_{ic} \tag{2}$$

where extinct_{ic} indicates whether language i indigenous to country c is extinct or not, MDI_i captures the Indigenous language's gender norms, and MDI_{ic}^{Dist} is the just-described distance in gender norms between Indigenous language i and colonizer c. $\mathbf{X}_{ic}$ indicates the full set of language-level controls discussed in section 4. This specification allows us to compare the importance of the language's *absolute* level of male dominance in explaining extinction to that of the *distance* from the colonizer.

To further explore the nature of the cultural clash, we then split the absolute distance measure into two components and separately assess the role of (i) the Indigenous language being *more male dominant* than the colonizer and (ii) the Indigenous language being *less male dominant* (i.e., more gender equal) than the colonizer:

$$\text{Language More MDI}_{ic} = \begin{cases} |\text{MDI}_i - \text{MDI}_c| & \text{if MDI}_i > \text{MDI}_c \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Language Less MDI}_{ic} = \begin{cases} |\text{MDI}_i - \text{MDI}_c| & \text{if MDI}_i < \text{MDI}_c \\ 0 & \text{otherwise.} \end{cases}$$

To tease out the separate effect of these two distinct components of cultural distance, we estimate the following:

$$\text{extinct}_{ic} = \omega \text{MDI}_i + \eta_1 \text{Language More MDI}_{ic} + \eta_2 \text{Language Less MDI}_{ic} + \rho' \mathbf{X}_{ic} + \epsilon_{ic} \tag{3}$$

Next, we re-run the same specification controlling for the colonizer's MDI (MDI_c), rather than the language's MDI:

$$\text{extinct}_{ic} = \mu \text{MDI}_c + \eta_1 \text{Language More MDI}_{ic} + \eta_2 \text{Language Less MDI}_{ic} + \rho' \mathbf{X}_{ic} + \epsilon_{ic} \tag{4}$$

Alternatively, to isolate the role of cultural distance from that of the colonizer's characteristics, we include colonizer fixed effectsoh, (τ_c) and thus leverage variation in cultural

distance among languages exposed to the same colonizer:

$$\text{extinct}_{ic} = \eta_1 \text{Language More MDI}_{ic} + \eta_2 \text{Language Less MDI}_{ic} + \tau_c + \rho' \mathbf{X}_{ic} + \epsilon_{ic} \quad (5)$$

In all specifications, we cluster the standard errors at the country level.

5.2 Results

Tables 3 and 4 present the results of these exercises. As can be seen in column 1 of Table 3, there remains a strong and significant relationship between a language's level of male dominance and its likelihood of extinction in the sample of Indigenous languages affected by European colonization. However, once including the absolute distance in gender norms from the colonizer as a regressor in column 2, the estimate of θ (the coefficient on the language's MDI) drops in magnitude and loses statistical significance. The positive and significant estimate of δ indicates that the larger the cultural distance in gender norms from the colonizer, the higher an Indigenous language risk of extinction, suggesting that cultural distance is a stronger predictor of extinction than the language's absolute gender norms.

Columns 3 to 5 unpack this relationship, and separately assess the role of two distinct clashes: one where the Indigenous group displays more gender equal norms than the colonizer and vice versa. An interesting pattern emerges: a language's risk of extinction increases only whenever its gender norms are more egalitarian than those of the colonizer, but not vice-versa. In both columns 3 and 5, the coefficient on *Language More MDI* is either close to zero or negative, and never statistically significant. To the contrary, in columns 4 and 5, the coefficient on *Language Less MDI* is large in magnitude and statistically significant at the 5 percent level.

Similar patterns can be observed when controlling for the colonizer's MDI or colonizer fixed effects, as shown in Table 4. In column 1, we find a positive correlation between the colonizer's MDI and an Indigenous language's risk of extinction. However, this coefficient becomes small and insignificant once the absolute cultural distance in the MDI is included in the specification in column 2. Again, columns 3 to 6 illustrate that the only positive and significant coefficient is the one on *Language more gender equal*, which survives to the inclusion of colonizer fixed effect, which net out the effect of colonizers' characteristics in explaining the cultural extinction of the Indigenous people they colonized.

5.3 Robustness

Our main results are robust to various tests, displayed in Appendix Table A-9. Findings and inference are insensitive to only keeping the most exogenous controls (columns 1-2), to robust standard errors (columns 3-4) and standard errors clustered at the language level (columns 5-6), to an alternative matching procedure of the MDI (columns 7-8), to a broader definition of extinction (columns 9-10), and to removing from the sample those languages that have gone extinct before 1900 (columns 11-12).

5.4 Interpretation

The results shown thus far point to a gender-based explanation for cultural extinction: the larger the mismatch in gender norms between colonizers and the colonized, the higher the latter's risk of cultural extinction.

However, distance in gender norms may correlate with other relevant distances from the colonizer that may, themselves, be responsible for the extinction of certain Indigenous groups. First, gender norms are a significant component of culture, so cultural distance in gender norms may be linked to broader cultural differences. These broader differences could explain the observed effects, rather than gender norms alone. To test this, we construct a measure of linguistic distance, a well-established measure of cultural distance. Since colonizers' languages often belong to entirely different linguistic families than those of Indigenous societies, the cladistic linguistic distance measure commonly used in the literature is typically 1 for most colonized countries. To address this, we restrict our sample to the Indigenous languages of India, Bangladesh, and Pakistan—three colonized countries where linguistic distance from the colonizer varies. In these countries, 42 percent of languages belong to the Indo-European family, the same as the colonizer's language, English. When running a horse-race between linguistic distance and distance in gender norms, we find that gender norms remain the most powerful explanation for cultural extinction, as shown in Table 5.

Second, we examine distances in pre-colonial institutional arrangements. If more hierarchical and centralized groups have an advantage in conquest due to, for example, greater military capacity (Zhang and Mace, 2021), differences in political complexity could influence cultural extinction. In a study closer to our setting, Cornell and Kalt (1995) examines post-colonial self-governing institutions among Native American groups, finding that

mismatches between formal institutions and pre-colonial informal ones create barriers to economic development. Since differences in gender norms may correlate with disparities in pre-colonial institutions between language groups and their colonizers, our estimates may reflect the mechanisms highlighted by Zhang and Mace (2021) and Cornell and Kalt (1995). To test for this, we use the measure of political complexity from the Ethnographic Atlas, a widely used proxy for pre-colonial institutions and state capacity (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013; Mayshar *et al.*, 2022). In Table 6, we find that, on its own, distance in political complexity predicts cultural extinction. However, when distance in political complexity is included alongside with distance in gender norms, the coefficient on distance in political complexity decreases in both magnitude and significance, while our coefficient of interest remains largey unchanged.

In light of these findings, we conclude that the gender-based explanation for cultural extinction is the most consistent with the data. Why does a more pronounced cultural mismatch in gender norms between the colonizer and the Indigenous group result in a higher likelihood of cultural extinction? Paula Gunn Allen describes the physical and cultural genocide of Indigenous American groups as follows:

> The physical and cultural genocide of American Indian tribes is and was mostly about patriarchal fear of gynocracy. The Puritans particularly, but also the Catholic, Quaker, and other Christian missionaries, like their secular counterparts, could not tolerate peoples who allowed women to occupy prominent positions and decision-making capacity at every level of society. —Paula Gunn Allen (1992). *The Sacred Hoop: Recovering the Feminine in American Indian Traditions*, p.19.

Besides importing their institutions and technology, colonizers also imported their own male-dominant social and economic structure into the territories they colonized. As Theda Perdue (1998) illustrates, European colonizers, in their efforts to "civilize" the Cherokee people, imposed that men replaced women as farmers, undermined the principle of matrilineal principles, outlawed polygamy, and replaced the extended family or kin-based network with the nuclear family structure. Similarly, Allen (1992) describes how colonizers also undermined women's political power:

> Politics played an even greater role in the destruction of the Cherokee gynocracy. [...] Cherokee women had the power to decide the fate of captives,

> decisions that were made by vote of the Women's Council [...] During the longtime colonization of the Cherokee along the Atlantic seaboard, the British worked hard to lessen the power of women in Cherokee affairs. —Paula Gunn Allen (1992). *The Sacred Hoop: Recovering the Feminine in American Indian Traditions*, p. 61-62.

Similar patterns are described by Oyèrónkẹ́ Oyěwùmí (1997) in the context of colonial Nigeria:

> It was unthinkable for the colonial government to recognize female leaders among the peoples they colonized, such as the Yoruba. [...] The transformation of state power to male-gender power was accomplished at one level by the exclusion of women from state structures. This was in sharp contrast to Yoruba state organization, in which power was not gender-determined. — Oyèrónkè Oyěwùmí (1997). *Invention of women: Making an African sense of Western gender discourses*, p. 124-125.

This newly imported socio-economic structure profoundly reshaped the role of men and women in Indigenous societies, tearing down their social fabric and disrupting their traditional economies. Our findings indicate that the larger the mismatch between Indigenous groups' pre-colonial gender relations and those imported by the colonizers, the higher the likelihood of cultural extinction.

6 Conclusion

This paper examines the determinants of cultural extinction, measured by language extinction, with a particular focus on the role of deep-rooted gender norms. We assemble a novel dataset that contains information on the language status together with other characteristics of the language group, such as their gender norms, geo-climatic and historical characteristics, among others.

Our empirical analysis establishes the relevance of gender norms, measured by the male dominance index (MDI) (Guarnieri and Tur-Prats, 2023), as a strong and robust predictor of language extinction. To further identify the role of gender norms as a determinant of cultural extinction, we exploit a major historical shock: European colonization. In this

setting, we construct a dyadic dataset in which we relate each colonizer to the Indigenous groups they contacted, and show that the greater the disparity in gender norms—particularly when the colonizer was more male-dominant than the Indigenous group—the higher the likelihood of the Indigenous language extinction.

To assess whether other factors might better explain language extinction, we test alternative sources of cultural or institutional mismatch, including linguistic distance and differences in pre-colonial political complexity. Our results do not support these alternative explanations and consistently show that gender norms remain the key explanatory factor.

We present several narratives that suggest that European colonizers imposed their own social structures and gender roles to the Indigenous groups they contacted. As a consequence of these actions, women's role in governance and economic production diminished, and traditional family structures and social orders were disrupted. We argue that this might have played a crucial role in the extinction of Indigenous languages.

Historical events such as colonization and modern forces such as globalization and urbanization are not only redefining economic relations but also cultural interconnectedness at a global scale. While substantial progress has been made in examining the economic effects of these global forces (for instance, Acemoglu *et al.* (2001) for the economic effects of colonization), their implications for the world's cultural landscape remain largely underexplored. This study provides a first systematic investigation of the cultural consequences of European colonization. Exploring the cultural impacts of other forms of inter-group contact presents a promising avenue for future research.

References

Abramitzky, R., Boustan, L. and Eriksson, K. (2020). Do immigrants assimilate more slowly today than in the past? *American Economic Review: Insights*, **2** (1), 125–141.

Acemoglu, D., Johnson, S. and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, **91** (5), 1369–1401.

Ahlerup, P. and Olsson, O. (2012). The roots of ethnic diversity. *Journal of Economic Growth*, **17**, 71–102.

Alesina, A. and Giuliano, P. (2015). Culture and institutions. *Journal of economic literature*, **53** (4), 898–944.

Allen, P. G. (1992). *The sacred hoop: Recovering the feminine in American Indian traditions*. Beacon Press.

Amano, T., Sandel, B., Eager, H., Bulteau, E., Svenning, J.-C., Dalsgaard, B., Rahbek, C., Davies, R. G. and Sutherland, W. J. (2014). Global distribution and drivers of language extinction risk. *Proceedings of the Royal Society B: Biological Sciences*, **281** (1793), 20141574.

Austin, P. K. and Sallabank, J. (2011). *The Cambridge handbook of endangered languages*. Cambridge University Press.

Bahrami-Rad, D., Becker, A. and Henrich, J. (2021). Tabulated nonsense? testing the validity of the ethnographic atlas. *Economics Letters*, **204**, 109880.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A. and Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, **5**, 180214.

Becker, B. (2019). Introducing coldat: the colonial dates dataset.

Besley, T. and Reynal-Querol, M. (2014). The legacy of historical conflict: Evidence from africa. *American Political Science Review*, **108** (2), 319–336.

Bisin, A. and Verdier, T. (2000). "beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics*, **115** (3), 955–988.

— and — (2017). *On the joint evolution of culture and institutions*. Tech. rep., National Bureau of Economic Research.

Blouin, A. and Dyer, J. (2022). Diversity and the world's endangered languages.

Boyd, R. and Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago press.

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S. and Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature ecology & evolution*, **6** (2), 163–173.

Centre for Research on the Epidemiology of Disasters (2023). Em-dat: The international disaster database.

Cornell, S. and Kalt, J. P. (1995). Where does economic development really come from? constitutional rule among the contemporary sioux and apache. *Economic Inquiry*, **33** (3), 402–426.

Crystal, D. (2002). *Language death*. Cambridge University Press.

Diamond, J. (2011). *Collapse: how societies choose to fail or succeed: revised edition*. Penguin.

Eberhard, D. M., Simons, G. F. and Fennig, C. D. (2024). *Ethnologue: Languages of the World*. SIL International, 27th edn.

Fernihough, A., Colvin, C. L. and McLaughlin, E. (2024). Mind your language: Explaining the retreat of the irish language frontier.

Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*, vol. 76. Multilingual matters.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in us schools after world war i. *The Review of Economic Studies*, **87** (1), 204–239.

Gallup, J. L., Sachs, J. D. and Mellinger, A. D. (1999). Geography and economic development. *International regional science review*, **22** (2), 179–232.

Galor, O., Klemp, M. and Wainstock, D. C. (2024). *Roots of Cultural Diversity*. Tech. rep., IZA Discussion Papers.

— and Özak, Ö. (2016). The agricultural origins of time preference. *American economic review*, **106** (10), 3064–3103.

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

Giavazzi, F., Petkov, I. and Schiantarelli, F. (2019). Culture: Persistence and evolution. *Journal of Economic Growth*, **24**, 117–154.

Giuliano, P. and Nunn, N. (2021). Understanding cultural persistence and change. *The Review of Economic Studies*, **88** (4), 1541–1581.

Guarnieri, E. and Tur-Prats, A. (2023). Cultural distance and conflict-related sexual violence. *The Quarterly Journal of Economics*, **138** (3), 1817–1861.

Hall, R. E. and Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *The quarterly journal of economics*, **114** (1), 83–116.

Hammarström, H., Forkel, R., Haspelmath, M. and Bank, S. (2023). Glottolog 4.8.

Humphreys, M. (2005). Natural resources, conflict, and conflict resolution: Uncovering the mechanisms. *Journal of conflict resolution*, **49** (4), 508–537.

Kik, A., Adamec, M., Aikhenvald, A. Y., Bajzekova, J., Baro, N., Bowern, C., Colwell, R. K., Drozd, P., Duda, P., Ibalim, S. *et al.* (2021). Language and ethnobiological skills decline precipitously in papua new guinea, the world's most linguistically diverse nation. *Proceedings of the National Academy of Sciences*, **118** (22), e2100096118.

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R. *et al.* (2016). D-place: A global database of cultural, linguistic and environmental diversity. *PloS one*, **11** (7), e0158391.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E. and Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American Journal of Tropical Medicine and Hygiene*, **70** (5), 486–498.

Lewis, M. P. (2014). Ethnologue: Languages of the world. *(No Title)*.

— and Simons, G. F. (2017). *Sustaining Language Use*. SIL International.

Lugones, M. (2016). The coloniality of gender. *Feminisms in Movement*, p. 35.

Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G. and Ni, F. (2009). Global signatures and dynamical origins of the little ice age and medieval climate anomaly. *Science*, **326** (5957), 1256–1260.

Mayshar, J., Moav, O. and Pascali, L. (2022). The origin of the state: Land productivity or appropriability? *Journal of Political Economy*, **130** (4), 1091–1144.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *American Economic Review*, **102** (4), 1508–1539.

— and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, **81** (1), 113–152.

— and Xue, M. M. (2021). Folklore. *The quarterly journal of economics*, **136** (4), 1993–2046.

Miller, C. (2022). Historical Conflict Event Dataset.

Mufwene, S. (2002). Colonisation, globalisation, and the future of languages in the twenty-first century. *International Journal on Multicultural Societies*, **4** (2), 162–193.

Natural Earth (2010). 1:10m coastline.

Nettle, D. (1998). Explaining global patterns of language diversity. *Journal of anthropological archaeology*, **17** (4), 354–374.

— and Romaine, S. (2000). *Vanishing voices: The extinction of the world's languages*. Oxford: Oxford University Press.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, **94** (1), 20–36.

Oyěwùmí, O. (1997). *The invention of women: Making an African sense of western gender discourses*. U of Minnesota Press.

Perdue, T. (1998). *Cherokee women: Gender and culture change, 1700-1835*. U of Nebraska Press.

Ploeg, F. v. d. (2011). Natural resources: curse or blessing? *Journal of Economic literature*, **49** (2), 366–420.

Rodrik, D., Subramanian, A. and Trebbi, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of economic growth*, **9**, 131–165.

SIL International (2023). ISO 639-3: Codes for the Representation of Names of Languages. Accessed: 2023-08-10.

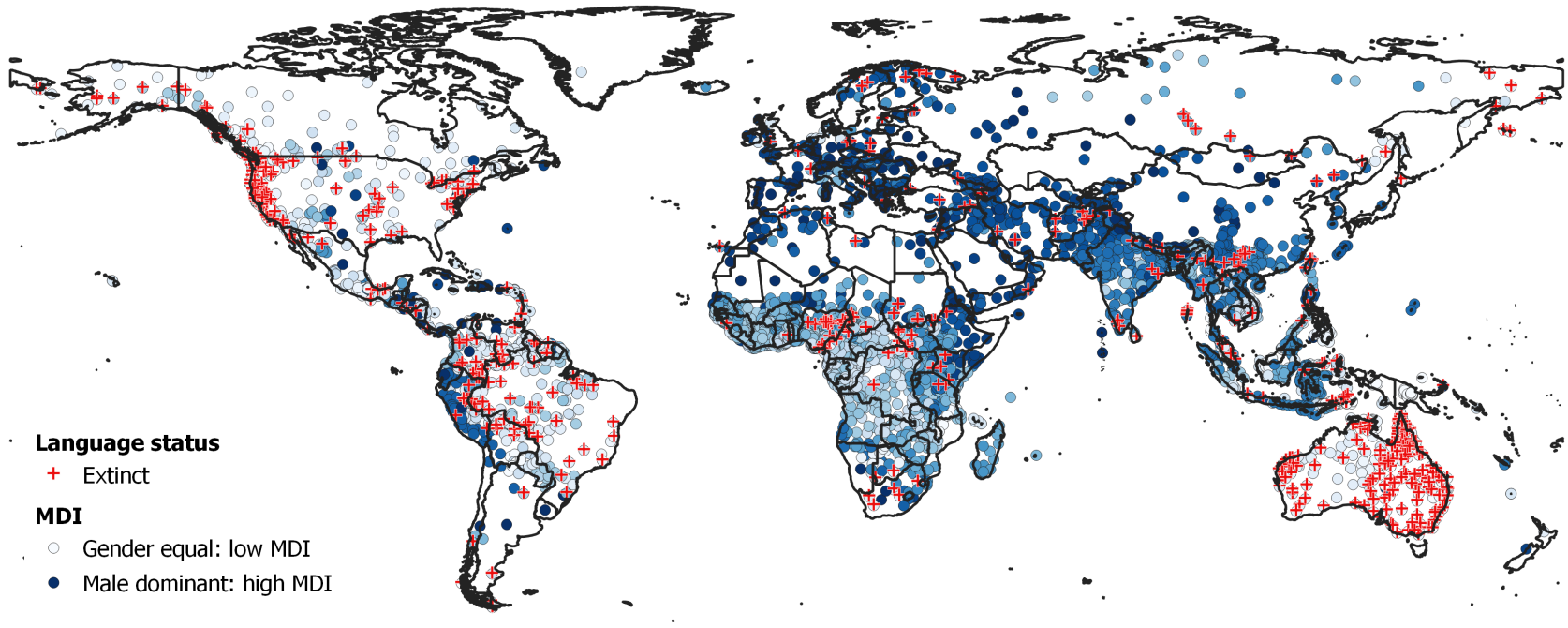
Tainter, J. (1988). *The collapse of complex societies*. Cambridge university press.

U.S. Geological Survey (1996). Gtopo30 global 30 arc-second elevation data set.

Zhang, H. and Mace, R. (2021). Cultural extinction in evolutionary perspective. *Evolutionary Human Sciences*, **3**, e30.

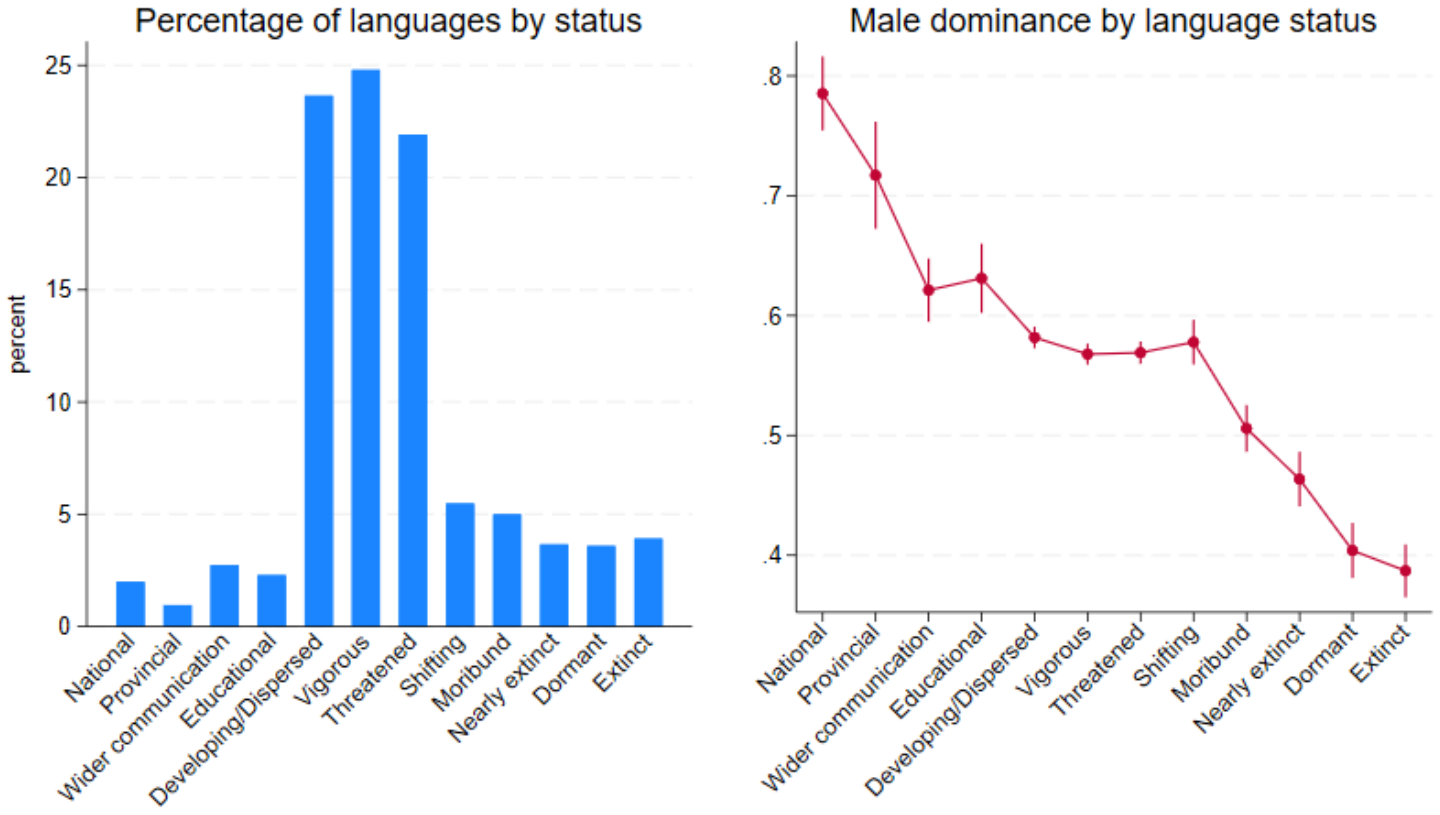
Figures and Tables

FIGURE 1: *Male Dominance and Language Extinction*



Notes: Each dot on the map represents a language. Darker shades of blue indicate higher degrees of male dominance of the society speaking the language based on Guarnieri & Tur-Prats' (2023) Male Dominance Index (MDI). Red crosses indicate languages that have gone extinct.

FIGURE 2: *Summary Statistics*



Notes: The histogram on the left displays the distribution of languages across categories in the EGDIS scale. The graph on the right illustrates the average male dominance of languages in each category of the EGDIS scale.

TABLE 1: *Male Dominance and Language Extinction*

	Dependent variable: extinct (0-1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MDI	-0.48**	-0.45***	-0.49***	-0.49***	-0.48***	-0.48***	-0.34***
	(0.228)	(0.156)	(0.148)	(0.146)	(0.146)	(0.148)	(0.127)
Mean dep. var.	0.107	0.107	0.107	0.107	0.107	0.107	0.107
Observations	5763	5763	5763	5763	5763	5763	5763
Clusters	172	172	172	172	172	172	172
Adjusted R^2	0.078	0.169	0.176	0.177	0.177	0.177	0.209
Geography		✓	✓	✓	✓	✓	✓
Climate and disease			✓	✓	✓	✓	✓
Natural resources				✓	✓	✓	✓
Natural disasters					✓	✓	✓
Exposure to war						✓	✓
Pre-colonial institutions							✓

Notes: The unit of observation is a language in a country. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Magnitude (column 7): a one standard deviation increase in the MDI decreases the likelihood of extinction by 0.2 standard deviations.

TABLE 2: *Male Dominance and Language Extinction within Countries*

	Dependent variable: extinct (0-1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MDI	-0.27***	-0.21**	-0.21**	-0.20**	-0.19**	-0.19**	-0.16*
	(0.092)	(0.086)	(0.085)	(0.084)	(0.083)	(0.083)	(0.084)
Mean dep. var.	0.107	0.107	0.107	0.107	0.107	0.107	0.107
Observations	5763	5763	5763	5763	5763	5763	5763
Clusters	172	172	172	172	172	172	172
Adjusted R^2	0.298	0.307	0.309	0.309	0.310	0.310	0.311
Geography		✓	✓	✓	✓	✓	✓
Climate and disease			✓	✓	✓	✓	✓
Natural resources				✓	✓	✓	✓
Natural disasters					✓	✓	✓
Exposure to war						✓	✓
Pre-colonial institutions							✓
Country fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is a language in a country. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. **Magnitude** (column 7): a one standard deviation increase in the MDI decreases the likelihood of extinction by 0.1 standard deviations.

TABLE 3: *Distance in Gender Norms from the Colonizer and Cultural Extinction (I)*

	Dependent variable: extinct (0-1)				
	(1)	(2)	(3)	(4)	(5)
Language's MDI	-0.31**	0.00	-0.31**	0.06	0.07
	(0.129)	(0.204)	(0.137)	(0.231)	(0.253)
Absolute distance in the MDI		0.32**			
		(0.153)			
Language more male dominant			0.06		-0.21
			(0.780)		(0.787)
Language more gender equal				0.37**	0.38**
				(0.173)	(0.183)
Mean dep. var.	0.113	0.113	0.113	0.113	0.113
Observations	4346	4346	4346	4346	4346
Clusters	106	106	106	106	106
Adjusted R^2	0.261	0.263	0.260	0.263	0.263
Language-level controls	✓	✓	✓	✓	✓

Notes: The unit of observation is a language in a country. The sample only includes countries that were colonized and only languages indigenous to the country. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 4: *Distance in Gender Norms from the Colonizer and Cultural Extinction (II)*

	Dependent variable: extinct (0-1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Colonizer's MDI	0.43*	0.08	0.41*	0.07	0.07	
	(0.219)	(0.255)	(0.220)	(0.254)	(0.253)	
Absolute distance in the MDI		0.31**				
		(0.135)				
Language more male dominant			-0.69		-0.14	-0.04
			(0.649)		(0.702)	(0.695)
Language more gender equal				0.31**	0.31**	0.29**
				(0.132)	(0.134)	(0.121)
Mean dep. var.	0.113	0.113	0.113	0.113	0.113	0.113
Observations	4346	4346	4346	4346	4346	4346
Clusters	106	106	106	106	106	106
Adjusted R^2	0.248	0.263	0.249	0.263	0.263	0.268
Language-level controls	✓	✓	✓	✓	✓	✓
Colonizer fixed effects						✓

Notes: The unit of observation is a language in a country. The sample only includes countries that were colonized and only languages indigenous to the country. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 5: *Interpretation: Distance in Gender Norms vs Linguistic Distance*

	Dependent variable: extinct (0-1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Language's MDI	-0.34**	-0.33**				
	(0.156)	(0.154)				
Language more male dominant			0.28		0.21	0.32
			(0.583)		(0.546)	(0.547)
Language more gender equal			0.40**		0.40**	0.40**
			(0.175)		(0.176)	(0.174)
Linguistic distance				0.03	0.04	0.03
				(0.053)	(0.047)	(0.046)
Mean dep. var.	0.038	0.038	0.038	0.038	0.038	0.038
Observations	446	446	446	446	446	446
Adjusted R^2	0.464	0.467	0.465	0.448	0.465	0.467
Language-level controls	✓	✓	✓	✓	✓	✓
Country fixed effects		✓				✓

Notes: The unit of observation is a language in a country. The sample only includes languages in India, Pakistan, and Bangladesh. Robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 6: *Interpretation: Distance in Gender Norms vs Distance in Institutions*

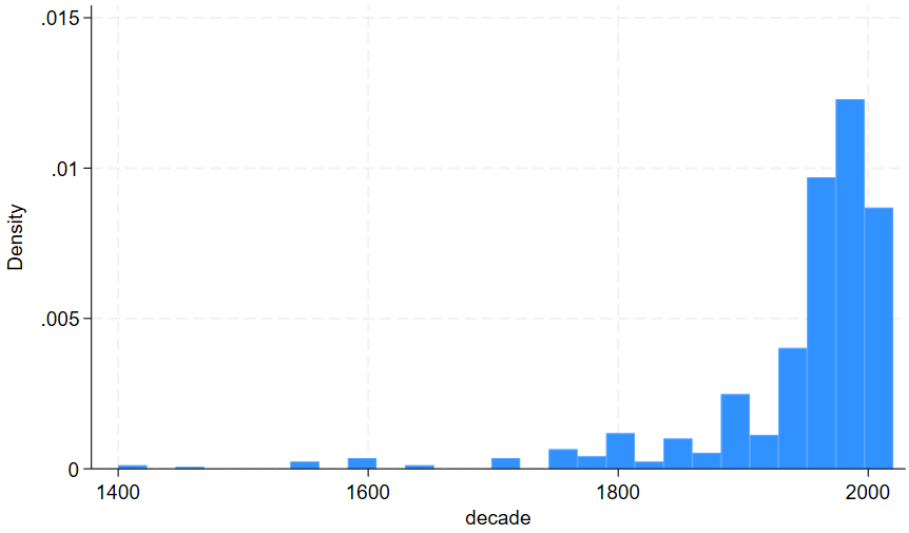
	Dependent variable: extinct (0-1)			
	(1)	(2)	(3)	(4)
Distance in political complexity	0.29***	0.14***		0.06*
	(0.110)	(0.039)		(0.030)
Language more male dominant			-0.14	-0.05
			(0.685)	(0.693)
Language more gender equal			0.32***	0.29**
			(0.122)	(0.123)
Mean dep. var.	0.113	0.113	0.113	0.113
Observations	4346	4346	4346	4346
Clusters	106	106	106	106
Adjusted R^2	0.088	0.255	0.268	0.267
Language-level controls		✓	✓	✓
Colonizer fixed effects	✓	✓	✓	✓

Notes: The unit of observation is a language in a country. The sample only includes countries that were colonized and only languages indigenous to the country. Language-level controls includes all the baseline controls, with the exception of political complexity. **Magnitudes** (column 4): a one standard deviation increase in distance in political complexity increases the likelihood of extinction by 0.04 standard deviations; a one standard deviation increase in a language being more gender equal than its colonizer increases the likelihood of extinction by 0.16 standard deviations. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix (Intended for Online Publication)

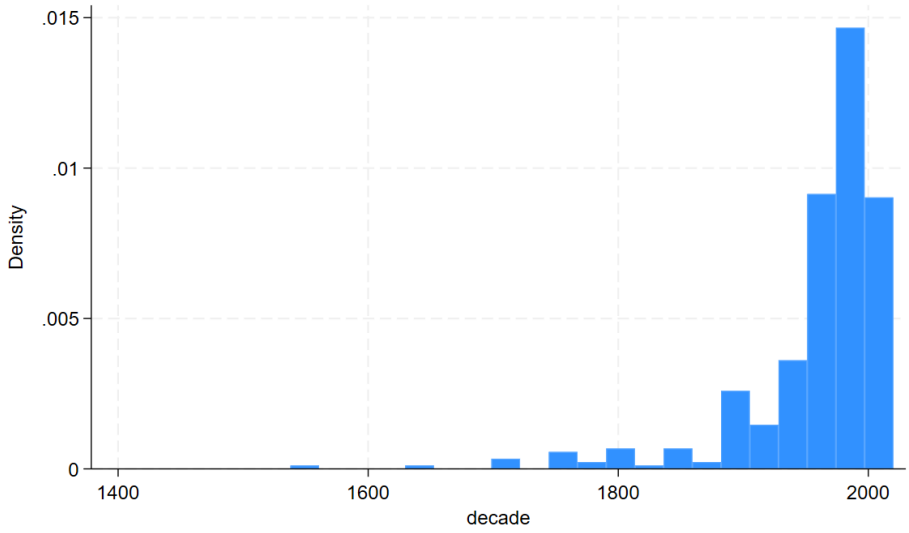
A Additional Figures and Tables

FIGURE A-1: *Language Extinction by Decade: Full Ethnologue Sample*



Notes: Based on information on 737 dormant or extinct language-countries (or 722 distinct languages) for which the Ethnologue provides information on the timing of extinction. Dates are approximate and based on when the last known speaker died.

FIGURE A-2: *Language Extinction by Decade: Estimation Sample*



Notes: Based on information on 386 dormant or extinct language-countries (or 383 distinct languages) in our estimation sample for which the Ethnologue provides information on the timing of extinction. Dates are approximate and based on when the last known speaker died.

TABLE A-1: *EGIDS scale*

Code	Status	Description
1	National	The language is used in education, work, mass media, and government at the national level.
2	Provincial	The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.
3	Wider Communication	The language is used in work and mass media without official status to transcend language differences across a region.
4	Educational	The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.
5	Developing	The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.
6a	Vigorous	The language is used for face-to-face communication by all generations and the situation is sustainable.
6b	Threatened	The language is used for face-to-face communication within all generations, but it is losing users.
7	Shifting	The child-bearing generation can use the language among themselves, but it is not being transmitted to children.
8a	Moribund	The only remaining active users of the language are members of the grandparent generation and older.
8b	Nearly Extinct	The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language.
9	Dormant	The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency.
10	Extinct	The language is no longer used, and no one retains a sense of ethnic identity associated with the language.

TABLE A-2: *Main Sample: Summary Statistics*

	Mean	St. dev.	Min	Max	N
Extinct	.1070623	.3092193	0	1	5763
MDI	.560714	.1807009	0	1	5763
Absolute latitude	17.53934	14.5855	.0013047	72.44424	5763
Caloric suitability index	1278.923	737.623	0	5047.091	5763
Distance to the coast	1113.916	672.842	.1857294	3633.362	5763
Elevation	678.8337	770.7886	-7.346774	5676.803	5763
Ruggedness	190.5475	214.5474	0	1606.767	5763
Climate variability	-.0343217	.2709756	-1.18	1.09	5763
Malaria index	7.121661	9.259626	0	37.54822	5763
Tropical climate	.5422033	.4761784	0	1	5763
Any natural resource	.4204407	.4936726	0	1	5763
Natural disasters (ln)	.747786	1.265559	0	8.445052	5763
Historical battles (ln)	.0941312	.4507724	0	5.932245	5763
Nomadic	.1248653	.3155232	0	1	5763
Political complexity	1.232525	1.189775	0	4	5763

TABLE A-3: *Colonization Sample: Summary Statistics*

	Mean	St. dev.	Min	Max	N
Extinct	.1134376	.3171635	0	1	4346
MDI	.5253912	.1730611	0	.955752	4346
Absolute distance MDI	.2696517	.1757715	.0009971	.8259152	4346
More male dominant	.0023496	.0129028	0	.2112644	4346
Less male dominant	.2673021	.178845	0	.8259152	4346
Absolute latitude	14.04495	11.92101	.0013047	69.87904	4346
Caloric suitability index	1202.884	710.0739	0	5047.091	4346
Distance to the coast	1086.542	682.759	.2632529	3633.362	4346
Elevation	604.3643	695.5915	-7.346774	5676.803	4346
Ruggedness	168.9336	199.0683	0	1606.767	4346
Climate variability	-.035585	.2627683	-1.18	1.09	4346
Malaria index	8.511525	9.663244	0	37.54822	4346
Tropical climate	.6343751	.4561198	0	1	4346
Any natural resource	.4077312	.4914693	0	1	4346
Natural disasters (ln)	.6738707	1.173431	0	6.932448	4346
Historical battles (ln)	.0387038	.259327	0	3.7612	4346
Nomadic	.1235324	.31807	0	1	4346
Political complexity	1.050394	1.087018	0	4	4346

TABLE A-4: *Horserace: MDI vs Single Ethnic Traits*

	Dependent variable: extinct (0-1)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MDI	-0.50**	-0.50***	-0.58***	-0.57***	-0.54***	-0.49**	-0.20**	-0.48**	-0.81***	-0.59***
	(0.220)	(0.164)	(0.196)	(0.187)	(0.184)	(0.202)	(0.103)	(0.207)	(0.204)	(0.191)
Mean dep. var.	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107	0.107
Control:	*plough*	*brideprice*	*polygyny*	*nuclear*	*matrilineal*	*fishing*	*gathering*	*pastoralism*	*non-herding husbandry*	*shifting agriculture*
Observations	5763	5763	5763	5763	5763	5763	5763	5763	5763	5763
Adjusted R^2	0.175	0.190	0.178	0.176	0.175	0.176	0.238	0.175	0.198	0.190

Notes: The unit of observation is a language in a country. Each column includes pre-determined language-level controls (geography, climate, natural resources). Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A-5: *Male Dominance Bias (Michalopoulos and Xue, 2021) and MDI*

	MDI		
	(1)	(2)	(3)
Male Dominance Bias	0.75***	0.35***	0.15**
	(0.193)	(0.102)	(0.075)
Mean dep. var.	0.561	0.561	0.561
Observations	5666	5666	5666
Adjusted R^2	0.079	0.335	0.496
Geography, climate, resources		✓	✓
Disasters, war, pre-colonial institutions			✓

Notes: The unit of observation is a language in a country. Both the MDI and the Male Dominance Bias are normalized and range between 0 and 1. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A-6: *Male Dominance Bias and Cultural Extinction*

	Dependent variable: extinct (0-1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male Dominance Bias	-0.31	-0.48*	-0.51**	-0.50**	-0.50**	-0.49**	-0.31*
	(0.296)	(0.249)	(0.248)	(0.243)	(0.241)	(0.242)	(0.180)
Mean dep. var.	0.107	0.107	0.107	0.107	0.107	0.107	0.107
Observations	5666	5666	5666	5666	5666	5666	5666
Clusters	172	172	172	172	172	172	172
Adjusted R^2	0.004	0.132	0.133	0.135	0.138	0.139	0.196
Geography		✓	✓	✓	✓	✓	✓
Climate and disease			✓	✓	✓	✓	✓
Natural resources				✓	✓	✓	✓
Natural disasters					✓	✓	✓
Exposure to war						✓	✓
Pre-colonial institutions							✓

Notes: The unit of observation is a language in a country. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Magnitude (column 7): a one standard deviation increase in the MDI decreases the likelihood of extinction by 0.07 standard deviations.

TABLE A-7: *Robustness (I)*

	only Indigenous languages				only direct matches		year of observation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MDI	-0.49**	-0.34**			-0.59**	-0.51***	-0.48**	-0.32***
	(0.242)	(0.133)			(0.243)	(0.157)	(0.225)	(0.121)
Alt. MDI			-0.44**	-0.31**				
			(0.217)	(0.118)				
Mean dep. var.	0.109	0.109	0.107	0.107	0.090	0.090	0.107	0.107
Observations	5452	5452	5763	5763	1096	1096	5740	5740
Clusters	167	167	172	172	144	144	172	172
Adjusted R^2	0.081	0.219	0.074	0.208	0.140	0.250	0.087	0.216
Baseline controls		✓		✓		✓		✓
Year of observation							✓	✓

Notes: The unit of observation is a language in a country. Columns 1 and 2 only include indigenous languages of the country. Columns 3 and 4 employ an MDI using an alternative matching procedure. Columns 5 and 6 only include languages that can be directly matched to the EA via the D-PLACE database. Columns 7 and 8 control for the year in which the Murdock ethnicity was observed by the ethnographer. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A-8: *Robustness (II)*

Dependent variable:	dying (0-1)		fully extinct (0-1)		extinct (0-1)			
					extinct after 1900		language level dataset	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MDI	-0.34**	-0.23*	-0.21*	-0.12**	-0.46**	-0.33***	-0.50***	-0.32***
	(0.170)	(0.132)	(0.119)	(0.054)	(0.223)	(0.125)	(0.031)	(0.033)
Mean dep. var.	0.177	0.177	0.039	0.039	0.102	0.102	0.122	0.122
Observations	5763	5763	5763	5763	5727	5727	4750	4750
Clusters	172	172	172	172	172	172		
Adjusted R^2	0.025	0.145	0.038	0.108	0.076	0.199	0.079	0.228
Baseline controls		✓		✓		✓		✓

Notes: The unit of observation is a language in a country in columns 1-6 and a language in columns 7-8. Columns 1 and 2 use a dependent variable that equals 1 if a language is dying (either shifting, moribund, nearly extinct or dormant in the EGIDS scale). Columns 3 and 4 use a dependent variable that equals 1 if a language is fully extinct in the EGIDS scale. In columns 5 and 6 languages that have gone extinct before 1900 are removed from the sample. Columns 7 and 8 collapse the dataset at the language level. Robust standard errors are reported in parentheses, clustered at the country level in columns 1-6. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A-9: *Robustness (III)*

Robustness:	controls		robust se		cluster language		alt. MDI		alt. dep. var.		extinct post 1900	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Language's MDI	-0.01		0.02		0.02				0.17		0.04	
	(0.341)		(0.094)		(0.098)				(0.309)		(0.250)	
Language more male dominant	0.45	0.44	-0.20	-0.17	-0.20	-0.17			-0.89*	-0.72*	-0.16	-0.12
	(0.838)	(0.658)	(0.400)	(0.391)	(0.425)	(0.415)			(0.519)	(0.405)	(0.757)	(0.685)
Language more gender equal	0.52**	0.53**	0.37***	0.34***	0.37***	0.34***			0.41	0.25**	0.38**	0.34**
	(0.233)	(0.211)	(0.088)	(0.032)	(0.092)	(0.033)			(0.285)	(0.119)	(0.172)	(0.146)
Colonizer's MDI		-0.01		0.02		0.02		0.03		0.17		0.04
		(0.341)		(0.094)		(0.098)		(0.261)		(0.309)		(0.250)
Language's MDI (alt)							0.03					
							(0.261)					
Language more male dominant (alt)							-0.62	-0.59				
							(0.585)	(0.475)				
Language more gender equal (alt)							0.33*	0.30**				
							(0.183)	(0.140)				
Mean dep. var.	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.165	0.165	0.109	0.109
Observations	4346	4346	4346	4346	4346	4346	4346	4346	4346	4346	4321	4321
Clusters	106	106			3888	3888	106	106	106	106	106	106
Adjusted R^2	0.224	0.224	0.262	0.262	0.262	0.262	0.258	0.258	0.175	0.175	0.255	0.255
Geography	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Climate	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Exposure to war			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pre-colonial institutions			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is a language in a country. Columns 1 and 2 only keep pre-determined, exogenous controls. Columns 3 and 4 report robust standard errors. Columns 5 and 6 cluster standard errors at the langauge level. Columns 7 and 8 employ an alternative MDI. Columns 9 and 10 use a dependent variable that equals 1 if a language is dying (either shifting, moribund, nearly extinct or dormant in the EGIDS scale). In columns 11 and 12, languages that have gone extinct before 1900 are removed from the sample. Robust standard errors are reported in parentheses, clustered at the country level when not otherwise indicated. *** p<0.01, ** p<0.05, * p<0.1.

B Data Appendix

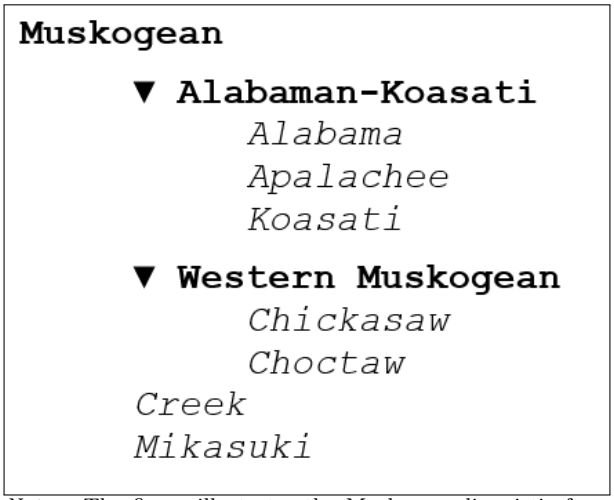
B.1 Procedure to Match Languages to Murdock Ethnic Groups

In this section, we describe the procedure we adopted to match Ethnologue languages to ethnic groups in the Murdock Ethnographic Atlas. The match is needed to assign a male dominance index to languages, our unit of observation in the analysis.

To achieve this, we draw on the methodology developed by Bahrami-Rad *et al.* (2021). As described in section 3, we start by employing the D-PLACE database (Kirby *et al.*, 2016), which maps 1,209 languages to 1,291 societies in the EA, to create what we refer to as "direct links" between languages and EA groups. Languages directly linked to an EA group receive the respective group's MDI.

To languages not directly linked to the EA, we assign the MDI of their nearest direct link(s) in the linguistic tree. Linguistic family trees provide genealogical classifications of languages, i.e., they trace how contemporary languages evolved from a common ancestor; languages sharing a common ancestor belong to the same linguistic family. We retrieve information on linguistic trees from the Glottolog database (Hammarström *et al.*, 2023). Figure A-3 illustrates the linguistic tree of the Muskogean family, which comprises 7 languages.[1] Languages, depicted in *italics* in the Figure, constitute the leaf nodes of the tree.

FIGURE A-3: *Example of Linguistic Tree*



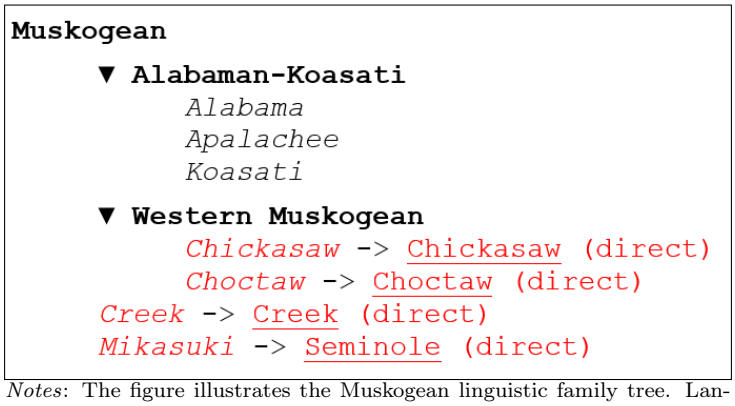
Notes: The figure illustrates the Muskogean linguistic family tree. Languages are the leaf nodes of the linguistic tree, and are indicated in *italics*. Source: `https://glottolog.org/resource/languoid/id/musk1252`

For some of these languages, the D-PLACE database provides us with a direct counterpart in the Murdock EA. Figure A-4 illustrates that four of the seven languages in

[1]Note that the typical family includes many more languages; the largest family is the Atlantic-Congo one, which embeds around 1,400 languages. In other cases, instead, languages have no genealogical relationship with other languages, and are therefore called "language isolates."

the Muskogean family (displayed in red) are directly linked to a Murdock ethnic group (displayed in red and underscored).

FIGURE A-4: *Direct Matches between Languages and Murdock Groups*

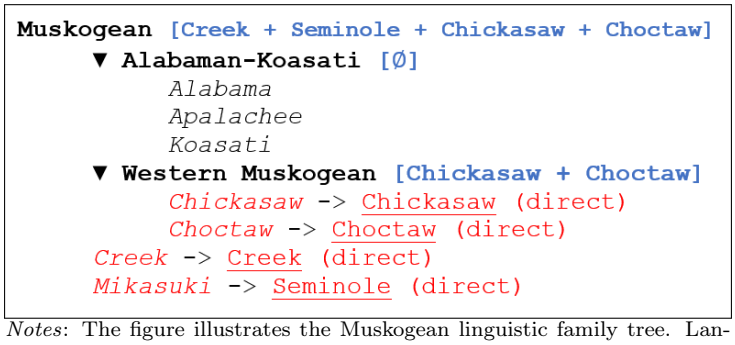


Notes: The figure illustrates the Muskogean linguistic family tree. Languages are the leaf nodes of the linguistic tree, and are indicated in *italics*. Languages that can be directly linked to a Murdock ethnic group are displayed in red. The name of the associated ethnicity is displayed in red and underscored. Sources: `https://glottolog.org/resource/languoid/id/musk1252` and D-PLACE database (Kirby *et al.*, 2016).

To match a Murdock group to the languages that remain unmatched, we rely on the structure of the linguistic tree and take advantage of the fact that languages that lie close in the tree are genealogically related and thus likely to be culturally similar one another. Based on this logic, to unmatched languages we assign the Murdock ethnic group associated to the nearest direct link(s) employing the following procedure.

To each node of the linguistic tree—depicted in bold—we assign all the Murdock groups directly linked to its child languages (if any). This procedure is illustrated in Figure A-5 below. As can be seen, the set of direct matches for the **Alabaman-Koasati** node is empty, as no child language is directly linked to the Murdock EA.

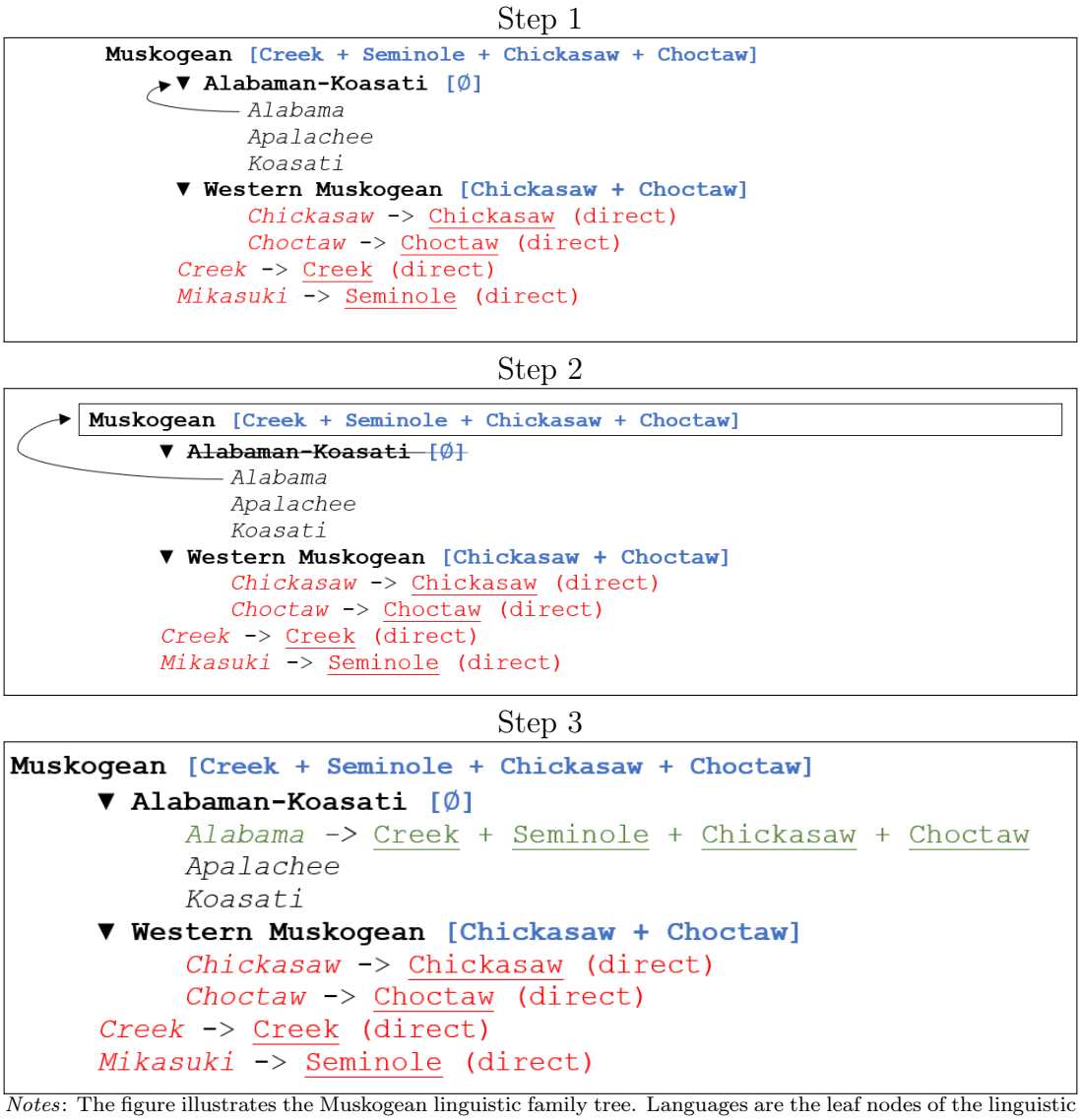
FIGURE A-5: *Assigning Murdock Groups to Nodes in the Linguistic Tree*



Notes: The figure illustrates the Muskogean linguistic family tree. Languages are the leaf nodes of the linguistic tree, and are indicated in *italics*. Languages that can be directly linked to a Murdock ethnic group are displayed in red. The name of the associated ethnicity is displayed in red and underscored. The Murdock groups assigned to each node are displayed in blue. Sources: `https://glottolog.org/resource/languoid/id/musk1252` and D-PLACE database (Kirby *et al.*, 2016).

Next, we take all unmatched languages and move up to each subsequent parent node in the tree, until a non-empty set of matches is found. Figure A-6 below illustrates the procedure for the *Alabama* language. We first move up to the **Alabaman-Kosati** node (step 1); because no child language has a direct Murdock match, we move up another node, the **Muskogean** one (step 2). As we encounter a non-empty set of matches, we assign four Murdock groups to the *Alabama* language: Creek, Seminole, Chickasaw, and Chocktaw (step 3).

FIGURE A-6: *Moving up the Tree to Identify Matches*

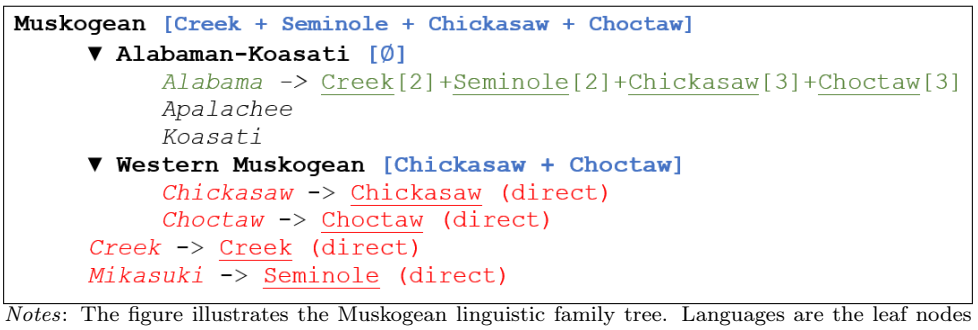


Notes: The figure illustrates the Muskogean linguistic family tree. Languages are the leaf nodes of the linguistic tree, and are indicated in *italics*. Languages that can be directly linked to a Murdock ethnic group are displayed in red. The name of the associated ethnicity is displayed in red and underscored. The Murdock groups assigned to each node are displayed in blue. The Murdock groups assigned to an unmatched languages are in green and underscored. Sources: `https://glottolog.org/resource/languoid/id/musk1252` and D-PLACE database (Kirby *et al.*, 2016).

As can be noted by observing the relationship between the *Alabama* language and the other languages in the family, some direct matches are "closer" to the *Alabama* language than others. We therefore assign a distance score to each match based on how many steps one needs to take in the linguistic tree to move from the unmatched language of interest (*Alabama*, in this case) to each matched language (*Chickasaw*, *Chocktaw*, *Creek*, and *Mikasuri*, in this case).

The respective scores are displayed in Figure A-7 below. To reach the *Creek* language, for instance, one needs to take two steps: move up to the **Alabaman-Koasati** node (1st step) and then horizontally to the *Creek* language (2nd step); the same applies to the Mikasuki language (and thus, the Seminole Murdock group). To reach the *Chickasaw* language, one needs to take three steps: move up to the **Alabaman-Koasati** node (1st step), horizontally to the *Western Muskogean* node (2nd step) and then down to the *Chickasaw* language (3rd step); the same applies to the *Choctaw* language.

FIGURE A-7: *Assigning a Distance Score to each Match*



Notes: The figure illustrates the Muskogean linguistic family tree. Languages are the leaf nodes of the linguistic tree, and are indicated in *italics*. Languages that can be directly linked to a Murdock ethnic group are displayed in red. The name of the associated ethnicity is displayed in red and underscored. The Murdock groups assigned to each node are displayed in blue. Sources: `https://glottolog.org/resource/languoid/id/musk1252` and D-PLACE database (Kirby *et al.*, 2016).

After establishing the Murdock matches and assigning a distance score to each of them, we assign to the *Alabama* language an MDI. We construct two alternative MDIs for languages like *Alabama*: a weighted average between the MDIs of the *Creek's*, *Seminole's*, *Chickasaw's*, and *Chocktaw's* MDIs, where the *Creek's* and *Seminole's* MDIs receive higher weights (this is our preferred measure); or a simple average using only the closest matches (in this case, *Creek's* and *Seminole's* MDIs only). The two measures are highly correlated.

B.2 Data Sources

Languages: Data on languages comes from the ISO 639-3 database. Every "distinct human language that has been documented, whether living, extinct, or constructed, and whether its modality is spoken, written or signed" (SIL International, 2023) is uniquely identified by a three-letter code, with a total of 8,552 entries in the database.
Source: `https://iso639-3.sil.org/`

Language status: Data on the current status of languages comes from the Ethnologue database (Eberhard *et al.*, 2024). Building on the ISO 639-3 classification, the Ethnologue assigns to each language spoken in a given country one out of 13 possible values of the Expanded Graded Intergenerational Disruption Scale (EGIDS) (Lewis and Simons, 2017). The higher the value in the EGIDS scale, the greater the disruption to the intergenerational transmission of the language. See Table A-3 for an illustration of the scale.
Source: `https://www.ethnologue.com/`

Male Dominance: Information on language group's gender norms comes from the male dominance index (MDI) developed by Guarnieri and Tur-Prats (2023), who exploit information from the Murdock Ethnographic Atlas, which describes cultural practices for more than 1,200 societies measured prior to industrialization and European contact. See Guarnieri and Tur-Prats's (2023) Online Appendix for details on the construction of the MDI and the EA variables included therein. For a description of the procedure we adopted to match the MDI to languages, please refer to Section B.1.
Sources: `https://doi.org/10.7910/DVN/B3LJGQ`

Latitude and Longitude: To match information on languages' location, we first use the Ethnologue map, which provides polygons for 7,626 living language-countries (or 6,425 distinct languages). We compute language's latitude and longitude by taking the centroid of each language's polygon. For the remaining languages not featured in the map we retrieve the point coordinates from the Glottolog (Hammarström *et al.*, 2023) and other sources such as the Endangered Languages Project or the language description provided by the Ethnologue.
Sources: Ethnologue Map, `https://www.endangeredlanguages.com`, `https://www.ethnologue.com/`

Elevation and Ruggedness: Data on elevation and ruggedness at the language level comes from combining the GTOPO30 rasters (U.S. Geological Survey, 1996) to the Ethnologue Map and languages' geographic coordinates. We compute the average elevation and ruggedness across each language polygon or a 0.25 degrees radius around the point location for cases in which information on the language's polygon is unavailable.
Source: `https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30`

Caloric Suitability: Data on caloric suitability at the language level comes from com-

bining the Caloric Suitability Index Developed by Galor and Özak (2016) to the Ethnologue Map and languages' geographic coordinates. We compute the average elevation and ruggedness across each language polygon or a 0.25 degrees radius around the point location for cases in which information on the language's polygon is unavailable.
Source: `https://ozak.github.io/Caloric-Suitability-Index/`

Distance to the Coast: We compute each language's distance to the coast using Natural Earth's (2010) coastline shapefiles combined with information on the location of languages.
Sources: `https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-coastline/`

Climate Variability: The measure of climate variability at the language level comes from Mann *et al.*'s (2009) data on the average annual temperature anomalies from 500 to 1900, which ranges from -1.54 to 1.1. We compute the average climate variability across each language polygon or a 0.25 degrees radius around the point location for cases in which information on the language's polygon is unavailable.
Source: `https://www.science.org/doi/10.1126/science.1177303`

Tropical Climate: Data on climate comes from the Köppen-Geiger map (Beck *et al.*, 2018). To obtain a language-level measure of tropical climate, we compute the ratio of a language's area (or 0.25 degrees radius around the point location for cases in which information on the language's polygon is unavailable) located in tropical climate.
Sources: `https://www.gloh2o.org/koppen/`

Natural Resources: Information on the presence of natural resources comes from the PRIO-grid dataset. We generate a binary indicator that equals to 1 if gold, petroleum, diamonds, or gems resources are present in a language polygon/radius and 0 otherwise.
Sources: `https://grid.prio.org/#/`

Disease Environment: Information on language's disease environment comes from Kiszewski *et al.*'s (2004) malaria index, ranging between 0 and 39. We compute the average malaria index across a language's area or across a 0.25 degrees radius around the point location for cases in which information on the language's polygon is unavailable.
Source: `https://pubmed.ncbi.nlm.nih.gov/15155980/`

Natural Disasters: Information on the occurrence of natural disasters comes from the EM-DAT database (Centre for Research on the Epidemiology of Disasters, 2023), which records the occurrence of natural disasters worldwide since 1900. We compute the number of natural disasters that occurred within each language polygon or 0.25 degrees radius.
Source: `https://www.emdat.be/`

Historical Conflict: Data on language's exposure to historical conflict comes from Miller's (2022) Historical Conflict Event Dataset. We compute the number of battles a linguistic group was exposed to from 1468 BC to the 18th century using language's

polygons/radius.
Sources: https://journals.sagepub.com/doi/suppl/10.1177/00220027221119085/suppl_file/sj-zip-2-jcr-10.1177_00220027221119085.zip

Settlement Patterns: Data on languages' settlement patterns (nomadic vs sedentary) comes from the Murdock EA, variable v30. For a description of the procedure we adopted to match the Murdock EA to languages, please refer to Section B.1.
Source: https://nathannunn.sites.olt.ubc.ca/files/2022/02/ethnographic_atlas_final.dta_.zip

Political Complexity: Data languages' pre-colonial political complexity (ranging from 0 to 4, from no political authority beyond the community to large states) comes from the Murdock EA, variable v33. For a description of the procedure we adopted to match the Murdock EA to languages, please refer to Section B.1.
Sources: https://nathannunn.sites.olt.ubc.ca/files/2022/02/ethnographic_atlas_final.dta_.zip

European Colonizer's Identity and Colonizer's Gender Norms: Data on the identity of the colonizer for each Indigenous language affected by European colonization comes from the Colonial Dates Dataset (COLDAT) dataset (Becker, 2019). We assign to each colonizer a measure of male dominance via the colonizer's language. Whenever a linguistic group was colonized by multiple colonial powers, we take the average MDI of all the colonizers of a given country.
Sources: https://doi.org/10.7910/DVN/T9SDEW

Male Bias: Data on an alternative measure of gender norms comes from Michalopoulos and Xue's (2021) male dominance bias (MDB) provided at the Murdock ethnic group level. For a description of the procedure we adopted to match the Murdock EA to languages, please refer to Section B.1.
Sources: https://doi.org/10.7910/DVN/IXOHKB