Male Dominance and Cultural Extinction*

Eleonora Guarnieri Ana Tur-Prats

March 26, 2026

Abstract: Why do some cultures go extinct while others prevail? In this paper, we show that a society's risk of cultural extinction—proxied by language loss—is shaped by its deeply rooted gender norms and by how these norms differ from those of competing cultures. Using the Ethnologue and the Male Dominance Index for a global sample of 4,750 languages, we document that languages from more gender-equal societies face a systematically higher likelihood of extinction than those from male-dominant societies. We propose a conceptual framework in which male-dominant societies are more expansionist and gender-egalitarian societies more permeable to outside influence, and test its central prediction in two empirical settings. Leveraging European colonization as a natural experiment, we find that Indigenous societies with more gender-equal norms than their colonizers are more vulnerable to cultural extinction, with historical narratives and evidence on language shift pointing to the forcible imposition of male-dominant structures and cultural assimilation as key mechanisms. Exploiting local variation in neighboring groups' gender norms across Africa, we uncover that gender-egalitarian societies face heightened cultural erosion when in contact with male-dominant neighbors. These findings provide the first global empirical evidence that gender relations are a central determinant of which cultures survive.

Keywords: Cultural diversity, Language extinction, Gender norms, Colonialism, Intergroup contact

JEL classification: Z10, O10, F54, J16

*Author affiliations and contacts: Guarnieri (University of Bristol, CEPR, and CESifo, eleonora.guarnieri@bristol.ac.uk), Tur-Prats (University of California at Merced and IZA, aturprats@ucmerced.edu). We thank numerous seminar and conference participants at Warwick, Queen Mary, ECARES, Luxembourg, St Gallen, Glasgow, Durham, Geneva, Paris Dauphine, University of Barcelona, Ca' Foscari University of Venice, Bristol, Birmingham, Rotterdam, UC Merced, York, UCLA Anderson, Frankfurt School/Goethe University, TU Dresden, BAMS Berlin, Applied Micro Workshop at Padua, the CEPR Annual Symposium, the Deep-Rooted Factors in Comparative Development conference at Brown, the Hayek Conference at LSE, the Authoritarianism and Nation Building workshop in Bergamo, the Yale/UB Workshop on Historical Political Economy, ASREC 2026, for useful comments and suggestions. Tiffany Ho, Johanna Lozano-Lavao, Zac Moghul, and Ntsima Ntsima provided excellent research assistance. All errors remain our own.

1 Introduction

Why do some cultures and their associated values go extinct while others prevail? Answering 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 a population, the collapse of a social group, or the abandonment of culture-specific practices (Boyd and Richerson, 1988; Zhang and Mace, 2021). These processes, in turn, may be driven by internal dynamics within a group, inter-group competition, or some combination of the two (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 and by how these norms differ from those of competing cultures.

Studying cultural extinction poses a fundamental empirical challenge: it is hard to determine whether and when a culture has disappeared. We address this challenge by using languages as the unit of observation and language extinction as a proxy for cultural extinction. Language is a primary carrier of culture, transmitted intergenerationally within populations, and often used to define cultural and ethnic boundaries (Ginsburgh and Weber, 2020). The loss of a language represents the disappearance of a fundamental, irreplaceable cultural asset. The scale of this loss in the modern era—the period examined in this paper—is unprecedented: nearly half of the world's approximately 7,000 languages are 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 a speaking population disappears entirely (as seen with certain Indigenous groups in the Americas), undergoes voluntary or forced assimilation into a dominant culture (such as Cornish speakers shifting to English), or faces some combination of these pressures (Nettle and Romaine, 2000).

The historical roles of men and women are a central feature of culture that reflects structural aspects of a society's economic and institutional organization. Our male dominance index (MDI) (Guarnieri and Tur-Prats, 2023) captures this concept, drawing on ten pre-industrial and pre-colonial ethnic characteristics from the Murdock Ethnographic Atlas (EA) relating to subsistence activities and family arrangements that have historically shaped the role of men and women in society and the economy. Our analysis is motivated by the striking empirical pattern shown in Figure 1: in the modern era, language extinction has been most prevalent among groups that were historically more gender equal, suggesting that gender structure may play a crucial yet overlooked role in the process of

cultural evolution.[1]

Why might gender structure influence a society's chances of survival? We propose a conceptual framework linking gender structure to patterns of cultural extinction. We argue that male-dominant societies are systematically more expansionist than their egalitarian counterparts, owing to two reinforcing features: the extraction of labor and resources that can be redirected toward military capacity and the demographic pressure generated by women's specialization in domestic and reproductive work.[2] Crucially, the success of expansion depends not only on the capacity of the expanding group, but also on the social structure of the target group: gender-egalitarian societies are more permeable to outside influence, as they rely on dispersed, consensus-based decisions, which can slow coordinated responses to external imminent aggression. These structural differences create asymmetries in capacity, and expansion will be more likely whenever male-dominant groups encounter more egalitarian groups. Once expansion takes place, cultural extinction arises as a result of biological replacement, the forcible imposition of new norms and institutions, voluntary assimilation, or a combination of these mechanisms. Taken together, these dynamics generate a testable prediction: cultural extinction should be systematically more likely among gender-egalitarian groups exposed to relatively more male-dominant competing cultures, a hypothesis we take to the data across two empirical settings: European colonization and cross-cultural variation across the African continent.

To this end, we assemble a novel dataset that combines information on languages and their status from the Ethnologue (Lewis, 2014), based on the Expanded Graded Intergenerational Disruption (EGIDS) Scale, with the gender structure 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 begin by testing whether gender structure is a strong predictor of cultural extinction.

[1]The focus of this paper is cultural extinction in the modern era of human cultural evolution. The Ethnologue provides information on the date of death of languages' last known speaker for a subset of the observations. Of the extinct and dormant languages in our sample for which this information is available, 97 percent saw their last speaker die after 1800, with the remaining 3 percent recording their last speaker between 1550 and 1790.

[2]We argue that male dominance—even though often operating hand in hand with other forms of hierarchical organization—is a stronger driver of expansionist tendencies than other forms of societal dominance alone: male dominance is self-reinforcing, less costly to sustain, and closely tied to women's specialization in domestic and reproductive work.

After regressing an indicator of whether or not a language is extinct on the MDI, we find that a one standard deviation increase in male dominance is associated with an 8.6 percentage points lower likelihood of cultural extinction. Notably, this relationship remains strong and significant after controlling for a wide large set of language-level geo-climatic characteristics and known historical determinants of economic development (Spolaore and Wacziarg, 2013). It also remains robust to the inclusion of country fixed effects. Taken together, these results point to male dominance as a strong and robust correlate of cultural survival.

A culture's survival is not determined in a vacuum. Cultural groups interact and compete, and through this contact some cultures perish while others endure. To test the central prediction of our framework—that gender-egalitarian groups are more likely to experience cultural extinction when facing male-dominant competing groups—we first 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.

We relate Indigenous languages to their colonizer in a dyadic framework using Becker's (2019) COLDAT dataset and find that a greater absolute distance in male dominance between the colonizer and the Indigenous group increases the latter's likelihood of extinction. Importantly, cultural distance in male dominance from the colonizer 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. Exploring this association further, we find that extinction is driven by a specific asymmetry: it is concentrated in cases where the Indigenous language is more gender equal than the colonizer, a finding that holds after the inclusion of colonizer fixed effects.

We then delve into the interpretation of our findings by considering the role of three competing explanations: group size at colonial contact, general cultural distance, and distance in other forms of societal dominance. First, we address the possibility that pre-contact group size confounds our results, as larger populations may have provided a demographic buffer against epidemic mortality and colonial pressure (Thornton, 1987). If more gender-equal groups were systematically smaller, this alone could explain our findings. We develop and validate a proxy for group size at contact by combining gridded

historical population estimates with the geographic distribution of (existing and extinct) Indigenous languages. We find that a group's size at contact is positively correlated with its male dominance and is itself a significant predictor of cultural survival. However, controlling for group size leaves the MDI distance coefficient largely unchanged.

Second, since male dominance is a core component of culture, our results may reflect general cultural incompatibility, and not incompatibility in gender structure specifically. To address this, we compute the linguistic distance between colonizer and Indigenous groups—a broadly-used proxy for cultural distance—and test whether it predicts cultural extinction. We do not find any significant effect. However, when both linguistic distance and male-dominance distance are included simultaneously, the latter retains its magnitude and statistical significance.

Third, since male dominance is one manifestation of a society's hierarchical nature, our results may reflect clashes in societal dominance rather than in gender structure specifically. We test this using two measures from the Ethnographic Atlas: political complexity, a proxy for pre-colonial institutions and state capacity (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013; Mayshar *et al.*, 2022); and a Societal Dominance Index (SDI), constructed as the first principal component of political complexity, history of slavery, and class stratification. Distance in both measures predicts extinction independently, but its effect attenuates substantially once distance in male dominance is accounted for, while male-dominance effects remain stable.

These results point toward a gender-based explanation for cultural extinction. To delve into mechanisms, we draw on historical narratives and theories from adjacent social sciences. European colonizers imported their male-dominant social and economic structure into the territories they colonized. As documented by Perdue (1998) for the Cherokee, this took concrete form: colonizers required men to replace women as farmers, undermined the principle of matrilineal kinship, and imposed the nuclear family structure. These disruptions are likely to have contributed to the breakdown of the most gender-equal Indigenous societies, eroding their social fabric and ultimately rendering them more vulnerable to cultural extinction.

To distinguish cultural assimilation from biological extinction, we use Ethnologue data on language shift, coding whether a language is shifting or has shifted to the colonizer's language or to a different language. This approach captures cases where speech communities survive physically but abandon their ancestral language. Our analysis shows that

Indigenous languages from less male-dominant societies were more likely to shift to the colonizer's language, whereas languages from more male-dominant societies were more resistant. The effect is specific to shifts toward the colonizer's language, suggesting that male-dominance asymmetries shaped susceptibility to colonial assimilation rather than general patterns of language change.

In a last step, we study inter-group contact at the local level in Africa. The availability of ethnolinguistic maps at two points in time—the pre-colonial period (Murdock map) and the present day (Ethnologue map)—allows us to construct an indicator at the grid-cell level capturing whether the ancestral language has survived up to the present. The continent's exceptional cultural diversity allows us to exploit local variation in the male dominance of neighboring groups. We find that the likelihood of language survival is significantly lower in cells neighboring male-dominant groups than in those neighboring gender-equal groups. Notably, the granularity of the data allows us to include homeland fixed effects, which net out group-specific factors, such as absolute defensive capacity or internal micro-level processes of cultural transmission. Importantly, this effect is asymmetric: cells of gender-egalitarian groups are disproportionately vulnerable to the pressures of male-dominant neighbors, whereas cells of male-dominant groups are largely unaffected by their neighbors' gender structure. These findings provide further evidence consistent with the prediction that gender-egalitarian groups face heightened risks of cultural erosion when in contact with male-dominant competing groups.

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). 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). We identify structural male-dominance as a key driver of language extinction and demonstrate that mismatches in gender structure with a dominant group place more gender-equal societies at heightened risk of linguistic extinction.

More broadly, our paper contributes to the literature on the deep historical roots of gender relations and their far-reaching consequences for social and economic life (Giuliano, 2018, 2020). Research in this tradition has shown that historical differences in subsistence

activities and kinship structures have shaped the division of labor, customs and norms around women's chastity, and competitive behavior along gender lines in ways that persist to this day (Alesina *et al.*, 2013; Becker, 2022; Gneezy *et al.*, 2009). Our MDI builds directly on this logic, and our findings extend this literature in a new direction: we show that the degree of male dominance not only shapes economic and social outcomes within societies, but also determines which societies survive.

Our paper also speaks to a growing body of work on cultural transmission and evolution. Research in economics builds on models of social learning and cultural socialization to study how norms and values emerge, persist, and change (Boyd and Richerson, 1985; Bisin and Verdier, 2001, 2025; Henrich, 2016). This has generated a rich literature examining 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), the horizontal diffusion of specific cultural traits, such as gender norms (Miho *et al.*, 2024), and the determinants of intra-group cultural heterogeneity (Galor *et al.*, 2024). Evolutionary anthropology, however, underscores that cultural evolution cannot be understood without accounting for cultural loss (Boyd and Richerson, 1985; Henrich, 2004). Cultural group selection theory, in particular, raises fundamental questions about how distinct groups fare when they compete (Henrich, 2004, 2016). While these questions have been explored theoretically, we provide the first global empirical analysis of the *disappearance* of established cultural entities and the conditions under which they go extinct. Our findings carry far-reaching implications for understanding the forces shaping contemporary levels of ethnolinguistic diversity and the intergenerational transmission of cultural traits.

Finally, our paper contributes 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; Dell, 2010; Michalopoulos and Papaioannou, 2016), its influence on the world's cultural landscape remains far less understood. We show that the cultural survival of Indigenous groups was shaped by the extent to which their level of male dominance diverged from that of the colonizing culture. 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

2.1 Societal Dominance and Expansionism

At its core, male dominance is a social arrangement in which one group (men) exercises power over another group (women), enabling the extraction of labor and resources (Lerner, 1986). As such, male dominance is one of the possible manifestations of a society's hierarchical nature. Comparable social structures can manifest through other forms of dominance, where power is concentrated in the hands of a select few through institutionalized political hierarchies, class stratification, or slavery, for instance. These forms of dominance are not mutually exclusive and often co-occur: more male-dominant societies tend to exhibit dominance in other domains as well, and vice-versa (Lerner, 1986).

Regardless of their specific form, we argue that these structures of societal dominance shape both a group's capacity for expansion, as well as its permeability to inter-group cultural competition. Hierarchical and extractive societies develop *expansionist capacity* by generating surpluses from the extraction of resources and labor, which can be redirected toward military development, lowering the costs of expansion and making territorial domination feasible (Spencer, 2010; Turchin *et al.*, 2013). In contrast, egalitarian groups may be especially *permeable to outside influence* because, unlike hierarchical societies, they often rely on dispersed, consensus-based decision making processes, which can slow the mobilization of coordinated responses to external threats (Boehm *et al.*, 1993). These structural differences create asymmetries in organizational capacity, such as the ability to mobilize populations, enforce decisions, and project authority beyond local networks. In turn, this gives more dominant societies a consistent advantage, making expansion more likely whenever they encounter more egalitarian groups (Tilly, 1992).[3]

2.2 Male Dominance or Other Forms of Societal Dominance?

While the relative importance of male dominance compared to other forms of hierarchical organization is ultimately an empirical question, we propose two conceptual arguments for why male dominance may be a more powerful driver of expansionist tendencies than other forms of societal dominance.

[3]When describing the European experience, Tilly (1992) argues that "Men who controlled concentrated means of coercion [...] ordinarily tried to use them to extend the range of population and resources over which they wielded power. When they encountered no one with comparable control of coercion, they conquered; when they met rivals, they made war." (Tilly, 1992, p. 14).

First, *male dominance is self-reinforcing and less costly to sustain.* Gender roles and norms are among the most pervasive and persistent features of a culture, operating from the micro-level of the family up to the macro-level of social institutions. Because these norms are internalized from birth, women—the subordinate group in male-dominant societies—may come to reproduce the very behaviors and beliefs that sustain their subordination (Lerner, 1986). Compliance is therefore often self-enforcing, maintained through deeply ingrained norms rather than through external coercion (Bussey and Bandura, 1999). By contrast, subordinate groups in other hierarchical systems (e.g., enslaved populations) are less likely to have undergone early socialization into norms of submission, and may consequently be more prone to resist or revolt. Maintaining political hierarchies, slavery, and class stratification therefore requires coercive infrastructure and substantial institutional capacity (Weber, 1946; Acemoglu and Wolitzky, 2011), diverting a portion of extracted surplus toward internal control rather than outward expansion.

Second, *women's specialization in domestic and reproductive labor is more specifically tied to male dominance* than to political hierarchy in general. In male-dominant societies, women's domestic and reproductive work is socially normalized and internalized, enabling higher fertility to be sustained without extensive coercive oversight. This demographic dynamic generates population pressure (Boserup, 1970; Goody, 1976), which strengthens incentives for territorial expansion (Diamond, 1997).

2.3 Expansionism, Permeability, and Cultural Extinction

Why would cultural extinction arise from the expansionist tendencies described above? We argue that, once expansion takes place, male-dominant groups are likely to impose their own socio-economic structures on the cultures they encounter. They will do so for at least two reasons: to preserve the productive, reproductive, and extractive technologies that allowed them to grow and expand in the first place; and for identity reasons, insofar as a social order in which gender roles diverge sharply from their own will be perceived as threatening or illegitimate, generating pressure to reorganize it (Akerlof and Kranton, 2000).

Cultural extinction may then unfold through three channels. First, through a biological channel: the physical elimination of the conquered population through genocidal violence and demographic replacement (Allentoft *et al.*, 2024; Reich, 2018; Turchin *et al.*, 2013). Second, through cultural oppression: the forcible imposition of new norms, in-

stitutions, and social arrangements on the surviving population (Turchin *et al.*, 2013). Third, through voluntary assimilation: conquered populations may themselves abandon their prior culture and adopt that of the expanding group. This is consistent with the prestige-bias mechanism in cultural evolution, whereby individuals preferentially adopt the traits and practices of those perceived as successful or high-status, using others' success as a proxy for cultural quality (Boyd and Richerson, 1985; Henrich and Gil-White, 2001). While these channels differ in the degree of coercion they involve, they share a common consequence: the erosion of the cultural fabric of gender-egalitarian groups onto which male-dominant ones have expanded, ultimately leading to their cultural extinction.

2.4 Historical Accounts

The Yamnaya expansion around 3000 BCE offers a notable prehistoric illustration of the just-illustrated dynamics. Drawing on archaeological and linguistic evidence, Gimbutas (1979) argues that Yamnaya people from the Eurasian steppe—who were male dominant, hierarchical, and warlike—expanded into Europe and imposed a male-centered social order on the more egalitarian Neolithic societies they encountered. The migration itself is supported by ancient DNA evidence (Reich, 2018), which documents steppe male lineages rapidly coming to dominate many European populations.[4]

A more recent illustration is the role of gender structure in European colonization. Anthropological accounts document how the male-dominated social order of the colonizers clashed with the more egalitarian structures of many Indigenous societies they encountered. Colonial expansion systematically imposed male dominance through preferential male access to resources, private property, wage labor, and the nuclear family on societies where it had not previously existed. Buenaventura-Posso and Brown (1980) describe this process in the case of the Bari of Colombia.[5] Such disruptions likely eroded the social

[4]Whether the pre-existing Old European societies were as peaceful and matrilineal as Gimbutas claims remains debated among archaeologists. However, the contrast in social organization between the expanding steppe pastoralists and the sedentary Neolithic farmers they encountered—the former more hierarchical and male-dominant—is broadly consistent with the available evidence (Anthony, 2008).

[5]"When Western capitalistic systems expand upon egalitarian communal societies, their actions involve the introduction and imposition of features such as the structure for general male dominance over females based materially on preferential male access to strategic resources; private ownership of land and a money economy; overall inequality in the form of differential access to privileges and resources which lead to the formation of socioeconomic classes; and an emphasis on the nuclear family, supported by a wage-earning male, as the ideal domestic unit. Wedded to their own experience of social and economic stratification, as well as to their concept of sex roles and statuses, such Western societies are unable to understand or accept the realities of a truly egalitarian society. Thus, outside contact with the Bari and their neighbors

fabric of gender-egalitarian Indigenous groups, contributing to the process of cultural extinction, which is the focus of our empirical analysis.

Our framework predicts that cultural extinction will be more likely among gender egalitarian groups facing relatively more male-dominant competing groups. To test this prediction, we leverage two empirical settings: European colonization, a natural experiment of modern history that exposed the majority of non-European societies to a dominant competing culture; and variation in the male dominance of neighboring cultural groups across the African continent. Alongside male dominance, we also consider political hierarchies and a composite index of societal dominance—encompassing political hierarchies, class stratification, and slavery—to assess the extent to which male dominance is the societal arrangement most conducive to expansion, domination, and ultimately cultural extinction.

3 Data

To study the relationship between gender structure and cultural extinction, we build a novel dataset at the language-country level combining information on language vitality, male dominance, 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 ethnolinguistic identity among speakers, an appealing feature for the specific purpose of this paper.[6] 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.[7]

has been typically male-centered, elitist, and divisive" (Buenaventura-Posso and Brown, 1980).

[6]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.

[7]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

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).[8] The higher the value in the EGIDS scale, the 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 Male Dominance and Other Ethnic Characteristics

Male Dominance. 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.

(SIL International, 2023).

[8]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.

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. Mean MDI in our main sample is 0.56 with a standard deviation of 0.31.

Pre-Colonial Institutions. Information on historical societal characteristics comes from the Ethnographic Atlas (EA). We construct an indicator of settlement patterns (nomadic vs. sedentary) and a measure of pre-colonial political complexity (ranging from 0 to 4, from no political authority beyond the community to large states). Following the same procedure as for our MDI, we also construct a Societal Dominance Index (SDI) at the ethnic group level by taking the first principal component of political complexity, history of slavery, and class stratification. The first principal component loads positively on all three variables and explains 62% of the total variance, consistent with the interpretation that it captures the concentration of social power in the hands of a select few.[9] Mean SDI in our colonization sample is 0.32 with a standard deviation of 0.25.

To assign the MDI and other ethnic characteristics 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 value. To languages not directly linked to the EA, we assign the MDI or pre-colonial institutions 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 structure in explaining extinction, we enrich our dataset with a large set of language-level geo-climatic and historical covariates. Information on the

[9]The loadings for the first component are 0.6335 for political complexity, 0.6408 for class stratification, and 0.4336 for slavery.

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 Glottolog (Hammarström *et al.*, 2023) and other sources such as the Endangered Languages Project or the language descriptions provided by the Ethnologue.[10] Additional details on data sources and matching are described in the Appendix.

Geography and Caloric Suitability. We compute average elevation and ruggedness for each language's area using the GTOPO30 rasters (U.S. Geological Survey, 1996), distance to the coast using Natural Earth's (2010) coastline shapefiles, and average caloric suitability using Galor and Özak's (2016) index.

Climate, Disease Environment, Natural Resources, and Disasters. First, we build a measure of climate variability at the language level based on Mann *et al.*'s (2009) data on average annual temperature anomalies from 500 to 1900, which range from -1.54 to 1.1. Second, we compute the share of a language's area located in a 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 constructing a binary indicator equal to 1 if gold, petroleum, diamonds, or gemstones are present in a language's polygon or 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 each language group's exposure to such events.

Historical Conflict. 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 end of the 18th century.

3.4 Identity of the Colonizer

Data on the identity of the colonizer comes from the Colonial Dates Dataset (COLDAT) dataset (Becker, 2019) and is provided at the country level. This dataset allows us to match every Indigenous language to its European colonizer. We assign to each colonizer

[10]For cases in which information on the language's polygon is unavailable, we take a 0.25-degree radius around the point location to compute the covariate averages. For estimating historical population at the language level, we apply Voronoi polygons as described below.

an MDI and other institutional characteristics through their corresponding ethnic group in the Murdock EA.

3.5 Historical Population

Comprehensive data on the size of Indigenous groups at contact is not available at a global scale. To address this, we develop and validate a proxy for group size at contact. We combine gridded population estimates from the History Database of the Global Environment (HYDE) with the geographic distribution of Indigenous languages, including extinct ones, mapped via Voronoi polygons (see Figure A-3). HYDE provides gridded population estimates at the century level from 10,000 BCE to 1700 CE, and at the decade level thereafter. This allows us to match each group's population estimate to the period immediately preceding its first colonial contact date, which is provided at the country level by COLDAT.

To validate our population proxy, we leverage variable EA202 in D-PLACE, which captures ethnic group population in a given year for groups in the Murdock EA.[11] For comparability, we construct our Voronoi-HYDE estimate using HYDE population data for the same decade as the D-PLACE's variable. We then regress the D-PLACE measure on our Voronoi-HYDE measure across a sample of more than 500 Murdock groups for which both measures are available. The binned scatterplot in Figure A-4 illustrates a positive relationship between the two measures. Table A-14 confirms this in a regression framework: the coefficient on our estimated population measure is positive and highly significant even after accounting for continent or country fixed effects. Taken together, these findings validate the use of our Voronoi-HYDE measure as a proxy for Indigenous group size at the time of first colonial contact.

3.6 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

[11]The population year does not coincide with the society's year of observation (i.e., the focal year in the Murdock EA), nor does it necessarily correspond to the society's pre-colonial period.

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 and thus indicating the endpoint of language decline. 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.[12] As can be seen, the endpoint of decline of most languages occurred over the last two centuries.

4 Male Dominance and Cultural Extinction

4.1 Empirical Strategy

The first goal of our empirical analysis is to establish whether a society's degree of male dominance is an important correlate of its likelihood of cultural extinction. To achieve this, we estimate the following linear probability model:

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

where extinct_{ic} is an indicator equal to 1 if a language i in country c is nearly extinct, dormant, or extinct, and 0 otherwise; MDI_i is the Male Dominance Index, a continuous variable ranging from 0 to 1 capturing the degree of male dominance of language i.[13] $\mathbf{X}_{ic}$ 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 linking them to economic development (Spolaore and Wacziarg, 2013), 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 comparative development literature to account for geographic effects (see, e.g., Acemoglu *et al.* (2001); Michalopoulos (2012); Rodrik *et al.* (2004)). We also control for distance to the coast and ruggedness to proxy for trade access and isolation, factors that may underlie economic development (Nunn

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

[13] 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.

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 disease burden are linked to differential income levels and growth through their impact on productivity and 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 has been directly linked to a society's risk of collapse (Diamond, 2011; Tainter, 1988).

Exposure to historical conflict. To account for a language's exposure to historical conflict, we control for the log number of battles that occurred prior to 1800 in a language's polygon. Historical conflict shapes long-run development through its effects on state formation and institutional quality (Tilly, 1992; Besley and Reynal-Querol, 2014).[14]

Precolonial arrangements and institutions. Finally, to disentangle the role of male dominance 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 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)).[15]

We include these sets of controls sequentially, from the arguably most exogenous and

[14] As highlighted in our conceptual framework, conflict might also be an integral part of a society's expansionary dynamics. However, the variable we include here serves a different purpose: by controlling for battles occurred prior to 1800, we aim to account for the long-term legacies of historical conflict *exposure*, rather than to capture expansionist capacity itself.

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

predetermined (e.g., geography) to the more endogenous ones, i.e., those that are likely co-evolving with gender structure (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 structure and the likelihood of extinction, confirming the visual pattern observed in Figure 1: the more male dominant a society's deep-rooted gender structure, 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 male dominance, but also of confounding factors that may correlate with gender structure and economic development. For instance, it could be that gender-egalitarian relations 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 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 structure may co-evolve 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 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 structure 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 male dominance from that of national institutions, colonial history, or policies common to all languages in a country.

4.3 Robustness

Each of the individual ethnic characteristics that form the MDI may affect cultural extinction directly and not (only) through gender structure. To assess whether the effect is driven by individual ethnic traits rather than the MDI, we take two steps. First, in Table A-5, 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. Table A-6 displays a strong correlation between the MDB and our MDI, further validating our measure of gender structure.[16] In Table A-7, we find that our main results hold when using 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-8 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[17] (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 who may have observed 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-9 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

[16] Conveniently, Michalopoulos and Xue (2021) provide the MDB at the Murdock ethnic group level.
[17] Appendix B.1 describes this alternative MDI in detail.

(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 structure and cultural extinction, operating over and above the confounding effect of other societal-level factors that may matter for both the formation a society's gender structure and its cultural survival.

5 Distance in Male Dominance, Colonization, and Cultural Extinction

Having established male dominance as a strong correlate of cultural extinction, we next turn to testing the main prediction of our conceptual framework, i.e., that cultural extinction is more likely among gender egalitarian groups facing relatively more male-dominant competing groups. To this end, we exploit a natural experiment of modern history—European colonization—which 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); 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 understanding 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.[18] Table A-4 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 male dominance in explaining extinction. We compute the absolute difference in gender structure between each Indigenous language i and the colonizer of its respective country c as follows:[19]

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

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 structure, and MDI_{ic}^{Dist} is the just-described distance in male dominance 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:

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

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

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

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

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

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

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

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

Alternatively, to isolate the role of cultural distance from that of the colonizer's characteristics, we include colonizer fixed effects, (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 structure 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 male dominance 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 male dominance.

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 relations than the colonizer and vice versa. An interesting pattern emerges: a language's risk of extinction increases only whenever its gender structure is more egalitarian than that 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-13. 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).

6 Interpretation and Mechanisms

The results shown thus far point to a gender-based explanation for cultural extinction: the larger the mismatch in gender structure between colonizers and the colonized, the higher the latter's risk of cultural extinction. However, distance in male dominance may correlate with other relevant distances from the colonizer that may, themselves, be responsible for the extinction of certain Indigenous groups. In this section we explore alternative

competing explanations for our results, focusing on group size at colonial contact, and the role of other relevant cultural and institutional factors.

6.1 Group Size at Colonial Contact

Group size at colonial contact is considered an important determinant of societies' survival chances in the aftermath of European colonization. For instance, Thornton (1987) argues that larger pre-contact populations provided a demographic buffer against epidemic mortality. Therefore, group size at colonial contact may confound our estimates if more egalitarian groups were systematically smaller.

In Table 5 we explore the role of group size at colonial contact in shaping cultural survival. Column (1) shows that group size is positively correlated with our MDI. The negative and significant coefficients of columns (2) and (3) show that group size at colonial contact is also a significant determinant of extinction, a finding consistent with Thornton (1987)'s argument that larger populations provided a demographic buffer against colonial and epidemic exposure. Crucially, however, the relationship between male dominance and cultural extinction survives the inclusion of group size at contact as control. Columns (4)-(7) show that the MDI coefficient remains negative and statistically significant after controlling for group size, both with and without colonizer fixed effects. Finally, columns (8) and (9) show that distance in male dominance—i.e., a language being more gender equal than the colonizer—remains a powerful predictor of extinction even after controlling for group size. Taken together, these results suggest that the effect of distance in gender structure on extinction is not confounded by group size at contact.

6.2 Linguistic Distance

Gender relations are a significant component of culture, so cultural distance in male dominance may be linked to broader cultural differences. These broader differences could explain the observed effects, rather than gender structure 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 male dominance, we find that male dominance remains the most powerful explanation for cultural extinction, as shown in Table 6.

6.3 Pre-colonial Institutions

Second, we turn to pre-colonial institutional arrangements, examining whether distances in hierarchical organization and other forms of societal dominance and drive our results. We start by using 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 7, 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 male dominance, the coefficient on distance in political complexity decreases in both magnitude and significance, while our coefficient of interest remains largely unchanged.

6.4 Other forms of Societal Dominance

As an alternative to political complexity alone, we construct a Social Dominance Index (SDI) designed to more broadly capture the concentration of political and social power in the hands of select few. The SDI is the first principal component of three variables from the Ethnographic Atlas: pre-colonial political complexity, history of slavery, and class stratification. While political complexity captures the hierarchical organization of a polity and its state capacity, slavery and class stratification add the dimension of institutionalized social inequality and subordination of certain groups to others (Michalopoulos, 2012; Lagerlöf, 2009). Together, these three variables capture not just the organizational sophistication of a society, but the degree to which power is concentrated in the hands of a dominant group. Following the same procedure as with the MDI, we take the first principal component as our summary index, which allows for a direct comparison between the two indices.

Table 9 unpacks the role of societal dominance vis-à-vis that of male dominance in shaping cultural extinction. As expected, societal dominance and male dominance are positively correlated, as illustrated in column (1). Moreover, societal dominance is itself an important determinant of extinction: as can be seen in columns (2) and (3), the

more socially dominant a cultural group, the less likely its cultural extinction, even after netting out the role of geographical and historical factors. Importantly, a horse-race between the MDI and the SDI in columns (4) to (6) reveals that the relationship between male dominance and cultural extinction remains large and significant, both with and without colonizer fixed effects, while the size of the societal dominance coefficient more than halves.

We then move to analyzing clashes in societal dominance versus male dominance between Indigenous groups and the colonizers. Column (7) shows that the absolute distance in the SDI is a positive and significant predictor of cultural extinction. As with male dominance, this effect reflects an asymmetry: column (8) shows that it is Indigenous groups that were *less* socially dominant than their colonizers that face a significantly higher extinction risk, while groups that were *more* socially dominant than their colonizers do not. The coefficient on the latter is larger in magnitude but imprecisely estimated, likely reflecting the rarity of such cases in the data.[20] Column (9) enters distance in both societal dominance and male dominance jointly. The coefficient on Indigenous groups being less socially dominant attenuates, while the coefficient on Indigenous groups being more gender equal than their colonizers remains large and statistically significant.

Taken together, these tests suggest that societal and male dominance are both predictors of cultural extinction. However, the fact that male dominance retains its explanatory power even after accounting for other forms of societal dominance is consistent with the arguments outlined in Section 2.2, pointing to male dominance as a more powerful driver of expansionism and, ultimately, cultural extinction. In the following subsections, we draw on historical narratives and illustrate how colonizing groups imposed their socio-economic structures on the Indigenous groups they encountered. We also present empirical evidence consistent with processes of cultural assimilation, whereby Indigenous groups abandoned their own language and adopted that of the expanding group, supporting the interpretation that our results reflect a cultural-assimilation channel underlying extinction, and not only a purely biological one.

[20]Indeed, in our sample colonizers tend to be more socially dominant than the Indigenous groups they encountered.

6.5 Mechanisms: Historical Narratives

Why would a larger mismatch in gender structure between colonizers and Indigenous groups predict cultural extinction? To shed light on the underlying mechanisms, we begin by drawing on historical narratives that document the dynamics of European colonization. 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.

Colonizers did not only import institutions and technology, they also imposed their own male-dominant social structure, actively dismantling Indigenous gender arrangements that conflicted with theirs. 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.

These dynamics were not limited to the Americas. Writing on colonial Nigeria Oyèrónkẹ́ Oyěwùmí (1997) documents how the British systematically excluded women from state structures they had previously occupied, transforming what had been a gender-neutral system of political power into an explicitly male-dominant one:

> 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.

These accounts suggest a common mechanism: colonizers systematically dismantled Indigenous gender arrangements that deviated from their own, eroding the social and political foundations on which those cultures rested. The larger this mismatch, the more disruptive the encounter, and the higher the likelihood of cultural extinction.

6.6 Mechanisms: Language Shift

The *Language Use* field of the Ethnologue database records whether a language group has shifted or is shifting toward another language. We use this to provide indirect evidence on the mechanism of extinction: a recorded shift suggests the speech community survived physically but abandoned its ancestral language, pointing toward assimilation, which may or may not involve population decline.[21] We extract two shift indicators using regular expression parsing: *shifting to* captures ongoing shifts (e.g., *"Shifting to Tok Pisin [tpi], especially youth"*) and *shifted to* captures completed shifts (*"All shifted to English [eng]"*). For each language, we also extract the ISO 639-3 code of the target language, which we match against the language(s) of the colonial power to construct two binary variables: *shift to colonizer*, equal to one if any recorded shift target is the colonizer's language, and *shift to other*, equal to one if the language shifted to a non-colonizer language.

Among Indigenous languages in our colonization sample that exhibit shift, 41% shifted toward the colonizer's language, rising to 58% among extinct languages. The remaining 59% shifted toward non-colonizer languages, most commonly English beyond its formal colonial sphere (11%), the Malay-Indonesian cluster (22% combined), and African regional lingua francas including Hausa, Fulfulde, and Arabic varieties (14% combined), languages

[21] This field is available for 86% of languages and contains semi-structured text describing the sociolinguist behavior of each language, including patterns of bilingualism, domain use, and, crucially for our purposes, language shift. Absence of a shift record may reflect incomplete documentation rather than the absence of assimilation.

whose dominance reflects centuries of trade and demographic pressure independent of European colonialism.

Table 9 examines whether cultural distance in male dominance predicts the direction of the language shift. Columns (1)-(5) use *shift to colonizer* as the dependent variable and columns (6)-(7) use *shift to other*. The Indigenous language's MDI in columns (1)-(2) is insignificant, indicating that relative distance from the colonizer, rather than absolute male dominance, drives this measure of assimilation. Languages more male dominant than the colonizer are significantly less likely to shift toward the colonizer's language, while less male dominant languages are significantly more likely to do so. The pattern holds within colonizer once colonizer fixed effects are added in columns (3)-(4), and is robust to restricting to completed shifts in column (5). Columns (6)-(7) show no effect on shift toward non-colonizer languages, suggesting that the result is specific to colonial assimilation rather than a general tendency of less male-dominant societies to shift languages.[22]

7 Inter-group Contact beyond Colonization: Spatial Analysis in Africa

European colonization provides a powerful historical example of how encounters between male-dominant colonizers and more gender-egalitarian Indigenous groups often resulted in the latter's cultural extinction. In this section, we offer an additional test of the central prediction of our framework—that gender-egalitarian groups are more likely to experience cultural erosion when facing male-dominant competing groups—by analyzing inter-group encounters in a spatial framework. To do so, we focus on the African continent, which provides a particularly suitable setting due to its exceptional degree of cultural diversity. Importantly, the spatial distribution of cultural groups can be observed at two distinct points in time: the pre-industrial, pre-colonial period (as captured by the Murdock map at the end of the nineteenth century) and the present day (as reflected in the Ethnologue map). This allows us to map the evolution of the continent's spatial cultural landscape, and thus construct a measure of language survival and loss at at a highly localized geographical level.

Specifically, we ask whether a group's language is more likely to disappear in areas located in close proximity to male-dominant neighbors, relative to areas bordered by more

[22]Results are also stable when restricting to observations with language-use information (columns (2), (4), and (7)), alleviating concerns that selection into Ethnologue documentation drives the findings.

egalitarian groups. To test for this, we split the African continent in equally-sized grid cells[23], and overlay them with both the pre-industrial and the present-day geographical distribution of languages. As each polygon in the Murdock map constitutes an ethnic group, we assign to each Murdock group a language using the "direct links" provided by D-PLACE (see Section 3). These direct links between Murdock groups and languages allow us to assign a language to each grid cell based on the Murdock map. We then construct a binary indicator of language survival that equals one if the ancestral language is still spoken in a given grid cell—as determined by the Ethnologue map—and zero otherwise. Figure 2 zooms into West Africa to illustrate the variation in language survival across grid cells, alongside the boundaries of ancestral ethnic homelands from the Murdock map. The figure reveals spatial heterogeneity in language survival, both across and within ancestral ethnic territories.

To address the concern that this measure solely captures measurement error in the Murdock and/or Ethnologue map, in Table A-10 in the Appendix we validate the measure by correlating it with the contemporary language status information from the Ethnologue. At the grid-cell level, we find a strong negative correlation between the grid-cell indicator of language survival and the language-based extinction indicator from Ethnologue (column (1)). This correlation remains strong and statistically significant when controlling for country fixed effects (column (2)), and when using the full EGIDS scale as the outcome variable (column (3)). Columns (4) to (6) show that the same pattern holds when the data is collapsed to the Murdock-group-by-country level, and language status is regressed on the share of an ethnic homeland's grid cells where the language remains spoken. While measurement error cannot be fully ruled out, these strong correlations suggest that our spatial measure of cultural survival is effective in capturing the contemporary status of languages.

We begin by exploiting cross-group variation in survival to establish whether male dominance is also a correlate of survival in this spatial analysis, as in Section 4. Specifically, we test whether grid cells inhabited by more male-dominated ethnic groups are more likely to have retained the ancestral language. To this end, we estimate the following specification:

$$\text{survival}_{ig} = \beta \text{MDI}_g + \gamma' \mathbf{X}_i + \epsilon_{ig} \tag{6}$$

[23]We leverage the PRIOgrid dataset, which partitions the world into $0.5^\circ \times 0.5^\circ$ grids, equivalent to 55 km at the Equator.

where survival$_{ig}$ is the aforementioned binary indicator that equals 1 if the ancestral language of group g is still spoken today in cell i and 0 otherwise. X_i is the same vector of geographic and historical controls outlined in the previous section, constructed at the grid cell level.[24] We additionally include as controls the size of the group, measured by the log number of cells forming its homeland, the distance of each cell from the closest neighboring ethnic group, as well as an indicator for whether a cell is split by a country border.[25] Standard errors are clustered at the historical ethnic homeland level. Table 10 presents the results. As the systematically positive and significant estimate for β reveals, cells of more male dominant ethnic groups are more likely to have retained their language.

Next, we move to leveraging within-group variation in (i) language survival and (ii) exposure to male-dominant versus gender-equal neighbors. Combined, these two sources of variation allow us to move beyond internal group characteristics and isolate the role of neighboring groups in shaping a language's likelihood of survival. We assign to each grid cell the male dominance of the closest neighboring group.[26] Crucially, the within-group variation in both the language survival as well as the male dominance of the immediate neighbor across grid cells—illustrated in Figure 3 for the Dialonke ethnic group—allows us to include ancestral group fixed effects, thereby isolating the role of neighbors' male dominance—that is, inter-group contact—from any confounding internal characteristics of the group itself. We estimate the following model:

$$\text{survival}_{igjc} = \theta \text{MDI}_{ij} + \gamma' \mathbf{X}_i + \alpha_g + \delta_c + \epsilon_{igjc} \tag{7}$$

where survival$_{igjc}$ is a binary indicator that equals 1 if the ancestral language of group g in country c is still spoken today in cell i with closest neighboring group j, and 0 otherwise. MDI$_{ij}$ is the male dominance index of cell i's closest neighbor j; X_i is the same vector of geographic and historical controls outlined in the previous section, constructed at the grid cell level, additionally including the distance of each cell from the closest neighboring ethnic group and an indicator for whether a cell is split by a country border or not. α_g and δ_c indicate ethnic homeland and country fixed effects, respectively.[27] The estimate of θ

[24] Political complexity and nomadism vary at the ethnic group level, as does the MDI.

[25] Country borders in Africa were largely established during the Scramble for Africa, which occurred around the time represented by the Murdock map.

[26] If a cell borders multiple groups, we assign to it the MDI of the most male dominant neighbor. Results are similar using the average MDI of all neighboring groups instead.

[27] Group size, political complexity and nomadism vary at the ethnic group level, therefore they are not included as controls as they are absorbed by group fixed effects. Country fixed effects can be included

captures the extent to which the probability of language survival differs for cells closer to male dominant groups as compared to cells closer to gender egalitarian neighbors, keeping all the internal characteristics of a group fixed. Beyond average effects, we also examine heterogeneity in responses. Consistent with our conceptual framework, we hypothesize that gender egalitarian groups are more susceptible to male-dominant neighbors, whereas male-dominant groups are largely insensitive to the expansionary pressures exerted by similarly male-dominant neighbouring groups. To test this, we estimate equation 7 separately for cells belonging to groups with MDI values above and below the sample median.

Table 11 presents the results of estimating equation 7 across the full sample and by subgroups defined by whether a language group's MDI falls above or below the sample median. Columns (1)–(3) report average effects, with the MDI of the closest neighbor entering negatively and significantly across all specifications, suggesting that proximity to a more male-dominant neighboring group is associated with a lower probability of language survival on average. This result is robust to restricting the sample to cells within 50km of a neighboring group and to the inclusion of the SDI of the closest neighbor, which enters insignificantly throughout. The heterogeneous effects, reported in columns (4)–(6) and (7)–(9), reveal a striking asymmetry consistent with our hypothesis. For groups with above-median MDI—that is, more male-dominant groups—the coefficient on neighbor MDI is small and statistically indistinguishable from zero across all specifications, suggesting that male-dominant groups are largely insensitive to the expansionist tendencies of equally male-dominant neighbors. By contrast, for groups with below-median MDI—i.e., the more gender-egalitarian groups—the coefficient is negative, substantially larger in magnitude, and highly significant. This indicates that cells in gender-egalitarian groups exhibit a probability of language survival when situated adjacent to male-dominant neighbors, as compared to cells located close to a more gender egalitarian neighbor. This confirms the hypothesis that such groups are disproportionately susceptible to the pressures exerted by male-dominant neighbors.[28]

in addition to group fixed effects as the same ethnic group may be spanning multiple countries. We note that, insofar as ethnic groups are geographically small and homogeneous, some of the controls included in $\mathbf{X}_i$ may generate multicollinearity problems when included together with group fixed effects. For this reason, we always show specifications omitting these controls.

[28]Similar conclusions can be drawn when using the average MDI of all closest neighbors rather than the highest (Table A-11), and when restricting the sample to cells featuring only the languages originally present in the Murdock map (Table A-12).

8 Conclusion

This paper examines the determinants of cultural extinction, proxied by language loss, with a particular focus on the role of societal gender structure. We assemble a novel dataset combining information on language status with the degree of male dominance of the associated language group, alongside geo-climatic and historical characteristics, for a global sample of languages. Our empirical analysis establishes the relevance of gender structure as a strong and robust predictor of language extinction. Yet, a culture's survival is not determined in a vacuum: cultural groups interact, and through this contact some cultures perish while others endure. We hypothesize that gender structure shapes these dynamics, as male-dominant societies are more expansionist and gender-egalitarian societies more vulnerable to outside influence. As a result, cultural extinction should be more likely among gender-egalitarian groups exposed to relatively more male-dominant competing cultures.

We uncover compelling evidence in support of this hypothesis. Using European colonization as a natural experiment, we show that the greater the disparity in male dominance—particularly when the colonizer was more male-dominant than the Indigenous group—the higher the likelihood of Indigenous cultural extinction. Male dominance remains the key explanatory factor even after accounting for group size at colonial contact and alternative sources of cultural or institutional mismatch, including linguistic distance and differences in pre-colonial institutions and other forms of societal dominance. Historical narratives illustrate how European colonizers systematically imposed their male-dominant structures on the Indigenous societies they encountered, displacing women from governance and economic production, dismantling matrilineal kinship systems, and enforcing the nuclear family. We also find suggestive evidence of cultural assimilation rather than pure biological extinction, with gender-equal Indigenous societies disproportionately likely to abandon their ancestral language in favor of the colonizer's. Finally, exploiting local variation in the gender structure of neighboring groups across Africa, we uncover similar patterns as in the European colonization experiment: gender-egalitarian societies face heightened risks of cultural erosion when in close proximity with male-dominant neighbors.

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. Understanding which societies are most at risk of cultural extinction, and why, is therefore essential. This paper provides the first global empirical evidence that gender relations shape inter-group contact, and are central in determining which cultures survive.

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.

— and Wolitzky, A. (2011). The economics of labor coercion. *Econometrica*, **79** (2), 555–600.

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

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, **115** (3), 715–753.

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

—, — and Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The quarterly journal of economics*, **128** (2), 469–530.

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

Allentoft, M. E., Sikora, M., Fischer, A., Sjögren, K.-G., Ingason, A., Macleod, R., Rosengren, A., Schulz Paulsson, B., Jørkov, M. L. S., Novosolov, M. *et al.* (2024). 100 ancient genomes show repeated population turnovers in neolithic denmark. *Nature*, **625** (7994), 329–337.

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.

Anthony, D. W. (2008). *The horse, the wheel, and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world*. Princeton University Press.

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, A. (2022). On the origins of restricting women's promiscuity. *Review of Economic Studies, forthcoming*.

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 — (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, **97** (2), 298–319.

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

— and — (2025). *Economic Models of Cultural Transmission*. Working Paper 33928, National Bureau of Economic Research.

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

Boehm, C., Barclay, H. B., Dentan, R. K., Dupre, M.-C., Hill, J. D., Kent, S., Knauft, B. M., Otterbein, K. F. and Rayner, S. (1993). Egalitarian behavior and reverse dominance hierarchy [and comments and reply]. *Current Anthropology*, **34** (3), 227–254.

Boserup, E. (1970). *Woman's Role in Economic Development*. London: George Allen & Unwin.

Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

— and — (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.

Buenaventura-Posso, E. and Brown, S. E. (1980). Forced transition from egalitarianism to male dominance: The bari of columbia. *ETIENNE and LEACOCK (eds.), op. cit.*

Bussey, K. and Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological review*, **106** (4), 676.

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

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

Dell, M. (2010). The persistent effects of peru's mining mita. *Econometrica*, **78** (6), 1863–1903.

Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. New York: W. W. Norton & Company.

— (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.

Gimbutas, M. (1979). The three waves of kurgan people into old europe, 4500–2500 bc. *Archives suisses d'anthropologie générale*, **43** (2), 113–137.

Ginsburgh, V. and Weber, S. (2020). The economics of language. *Journal of Economic Literature*, **58** (2), 348–404.

Giuliano, P. (2018). Gender: An historical perspective. In S. L. Averett, L. M. Argys and S. D. Hoffman (eds.), *Oxford Handbook on the Economics of Women*, New York: Oxford University Press, pp. 645–672.

— (2020). Gender and culture. *Oxford Review of Economic Policy*, **36** (4), 944–961.

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

Gneezy, U., Leonard, K. L. and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, **77** (5), 1637–1664.

Goody, J. (1976). *Production and Reproduction: A Comparative Study of the Domestic Domain*. No. 17 in Cambridge Studies in Social Anthropology, Cambridge: Cambridge University Press.

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.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, **53** (1), 3–35.

— (2016). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press.

— and Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior*, **22** (3), 165–196.

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.

Lagerlöf, N.-P. (2009). Slavery and other property rights. *Review of Economic Studies*, **76** (1), 319–342.

Lerner, G. (1986). *The Creation of Patriarchy*. New York: Oxford University Press.

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 — (2016). The long-run effects of the scramble for africa. *American Economic Review*, **106** (7), 1802–1848.

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

Miho, A., Jarotschkin, A. and Zhuravskaya, E. (2024). Diffusion of gender norms: Evidence from stalin's ethnic deportations. *Journal of the European Economic Association*, **22** (2), 475–527.

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.

Reich, D. (2018). *Who we are and how we got here: Ancient DNA and the new science of the human past*. Oxford University Press.

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.

Spencer, C. S. (2010). Territorial expansion and primary state formation. *Proceedings of the National Academy of Sciences*, **107** (16), 7119–7126.

Spolaore, E. and Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of economic literature*, **51** (2), 325–369.

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

Thornton, R. (1987). *American Indian holocaust and survival: A population history since 1492*, vol. 186. University of Oklahoma Press.

Tilly, C. (1992). *Coercion, capital, and European states, AD 990-1992*, vol. 19. Blackwell Oxford.

Turchin, P., Currie, T. E., Turner, E. A. L. and Gavrilets, S. (2013). War, space, and the evolution of old world complex societies. *Proceedings of the National Academy of Sciences*, **110** (41), 16384–16389.

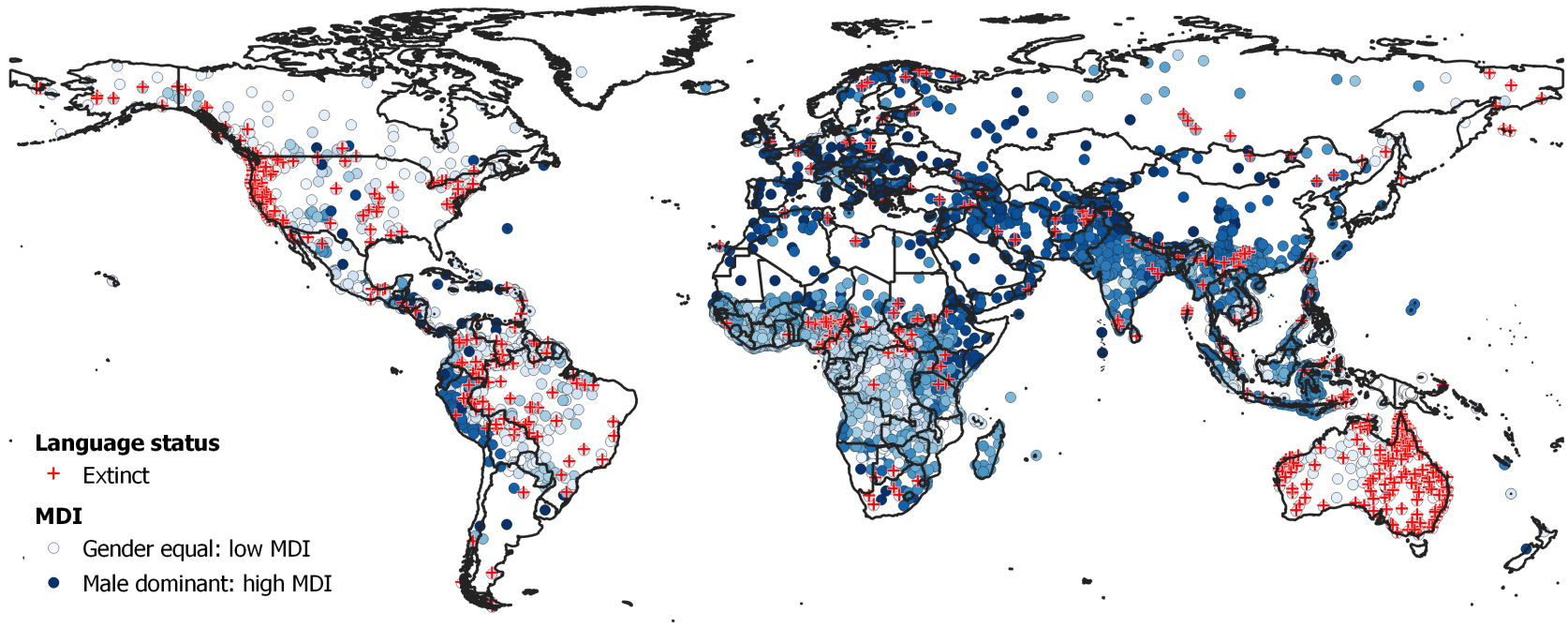
U.S. Geological Survey (1996). Gtopo30 global 30 arc-second elevation data set.

Weber, M. (1946). *From Max Weber: Essays in Sociology*. New York: Oxford University Press.

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

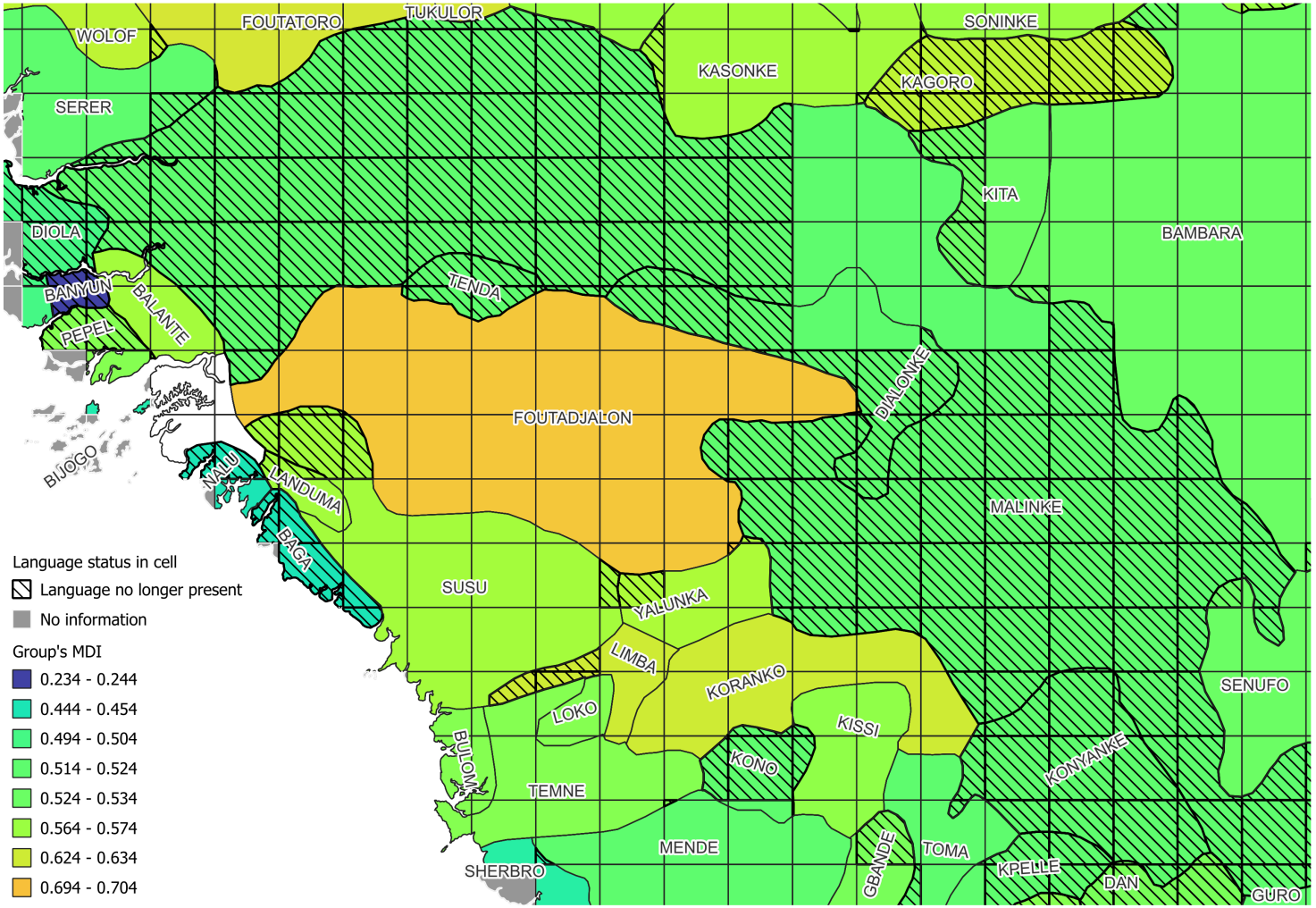
Figures and Tables

FIGURE 1: *Male Dominance and Language Extinction*



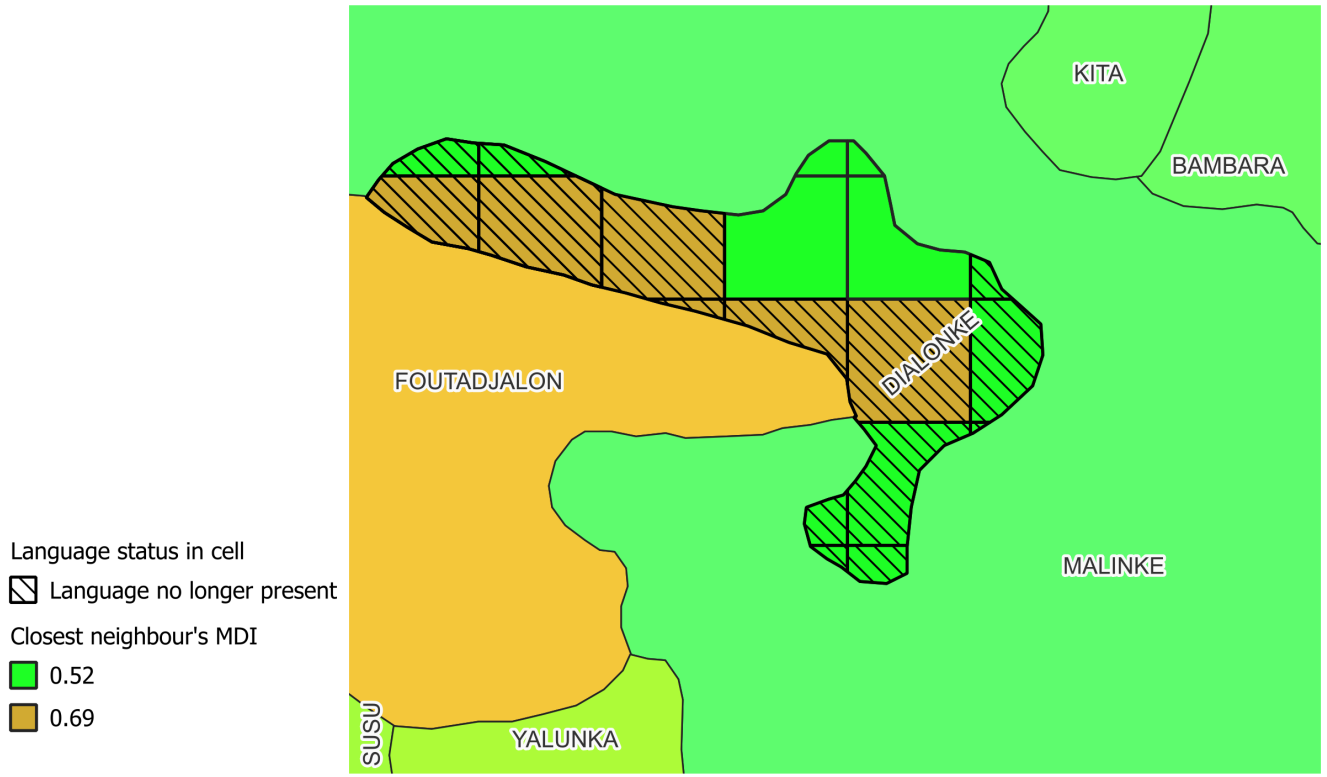
Notes: Each dot on the map represents a language. Darker shades of blue indicate higher degrees of male dominance of the society speaking the language based on Guarnieri & Tur-Prats' (2023) Male Dominance Index (MDI). Red crosses indicate languages that have gone extinct.

FIGURE 2: *Cultural Extinction in Space: Illustration from West Africa*



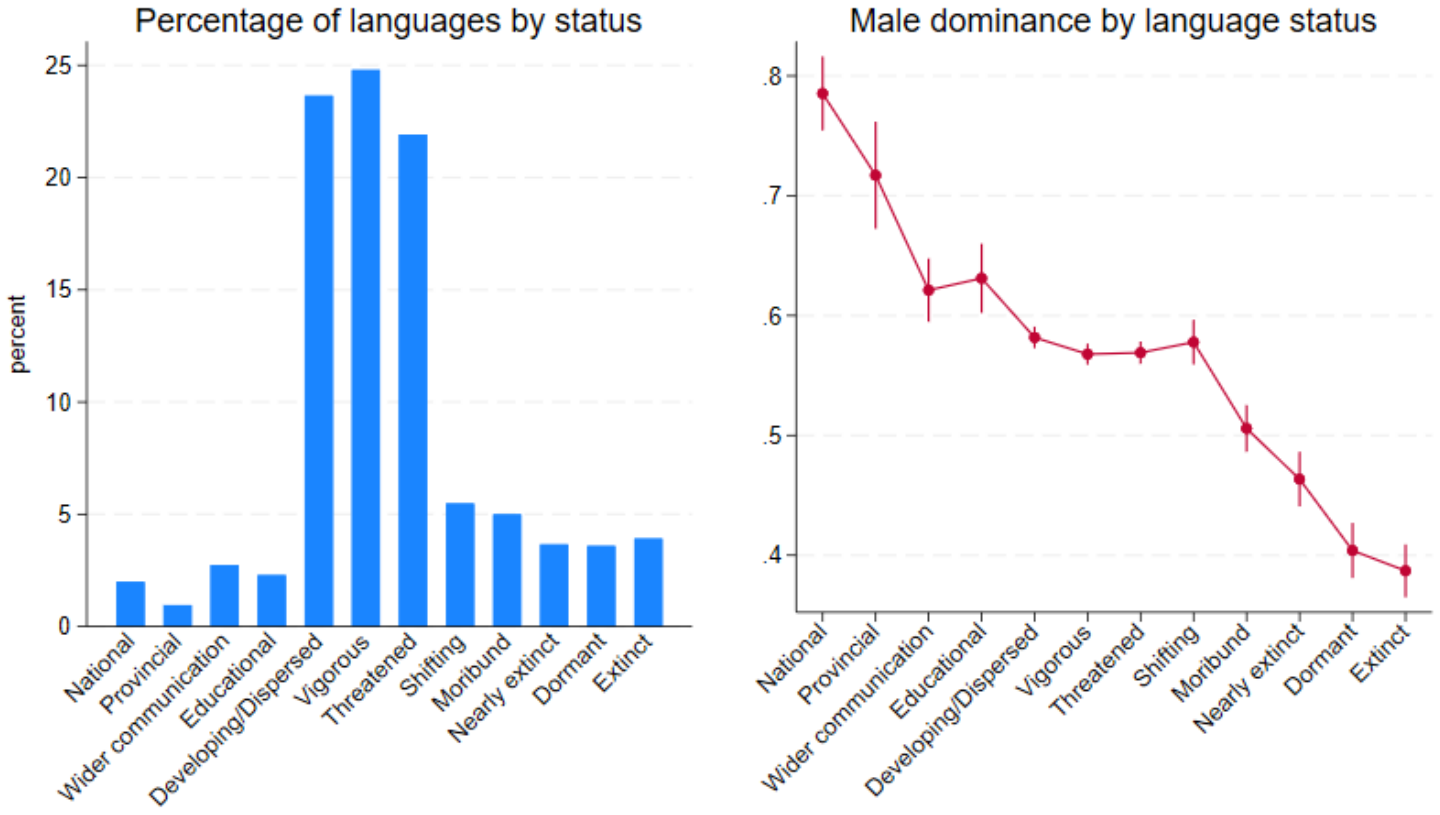
Notes: Ancestral Murdock ethnic homelands in West Africa, partitioned into 0.5° × 0.5° (∼55km) grid cells. Colors reflect each homeland's degree of male dominance, ranging from gender-equal (blue/green) to male-dominant (yellow/orange). Hatched grid cells indicate that the ancestral language is no longer spoken there today.

FIGURE 3: *Neighbouring Groups' Male Dominance and Cultural Extinction in Space: Identifying Variation*



Notes: The map displays the ancestral Murdock ethnic homeland of the Dialonke group, partitioned into 0.5° × 0.5° (∼55km) grid cells, along with its neighboring groups and their respective MDI. Cells are shaded by the degree of male dominance of the Dialonke's closest neighbor. Hatched grid cells indicate that the ancestral language is no longer spoken there today.

FIGURE 4: *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: *The Role of Group Size at Colonial Contact*

	Dependent variable:								
	MDI	extinct (0-1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln group population at contact	0.03***	-0.04**	-0.02***		-0.02***		-0.02***		-0.02***
	(0.007)	(0.018)	(0.008)		(0.006)		(0.006)		(0.006)
Language's MDI				-0.34**	-0.27**	-0.32**	-0.25**		
				(0.148)	(0.128)	(0.136)	(0.117)		
Language more male dominant								-0.05	-0.06
								(0.665)	(0.671)
Language more gender equal								0.31**	0.24**
								(0.136)	(0.117)
Mean dep. var.	0.527	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.111
Observations	4289	4289	4289	4289	4289	4289	4289	4289	4289
Clusters	106	106	106	106	106	106	106	106	106
Adjusted R^2	0.233	0.088	0.259	0.262	0.271	0.272	0.280	0.270	0.279
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 by country are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 6: *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 7: *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. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 8: *The Role of Societal Dominance*

	Dependent variable:								
	MDI	extinct (0-1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Language's SDI	0.39***	-0.32**	-0.25***		-0.11**	-0.11***			
	(0.078)	(0.152)	(0.089)		(0.043)	(0.038)			
Language's MDI				-0.49**	-0.41**	-0.37**			
				(0.196)	(0.181)	(0.161)			
Absolute distance in SDI							0.22***		
							(0.072)		
Language more socially dominant								0.43	0.47
								(0.288)	(0.282)
Language less socially dominant								0.29***	0.17***
								(0.104)	(0.060)
Language more male dominant									0.80
									(0.699)
Language more gender equal									0.38**
									(0.164)
Mean dep. var.	0.517	0.119	0.119	0.119	0.119	0.119	0.119	0.119	0.119
Observations	3959	3959	3959	3959	3959	3959	3959	3959	3959
Clusters	102	102	102	102	102	102	102	102	102
Adjusted R^2	0.307	0.061	0.203	0.228	0.232	0.245	0.221	0.224	0.248
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 other institutions. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 9: *Male Dominance and Language Shift*

	Dependent variable:						
	Shift to colonizer's language (0-1)					Shift to other language (0-1)	
					completed only		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Language's MDI	0.38	0.39					
	(0.290)	(0.350)					
Language more male dominant	-1.27**	-1.53**	-0.90**	-1.14**	-0.75*	-0.05	0.09
	(0.540)	(0.658)	(0.432)	(0.501)	(0.402)	(0.415)	(0.438)
Language less male dominant	0.62**	0.69**	0.22*	0.27*	0.21*	0.02	0.03
	(0.250)	(0.282)	(0.115)	(0.150)	(0.113)	(0.077)	(0.086)
Mean dep. var.	0.073	0.083	0.073	0.083	0.059	0.103	0.118
Observations	4346	3796	4346	3796	4346	4346	3796
Clusters	106	100	106	100	106	106	100
Adjusted R^2	0.240	0.316	0.254	0.333	0.254	0.056	0.071
Language Controls	✓	✓	✓	✓	✓	✓	✓
Colonizer FEs			✓	✓	✓	✓	✓
Language Use Information		✓		✓			✓

Notes: The unit of observation is a language in a country. Information on language use is at the language level. The sample only includes countries that were colonized and only languages indigenous to the country. *Language Use Information* restricts the sample to those observations for which the Ethnologue included language use information (88% of the regression sample). Column (5) uses a restricted dependent variable defined only over completed language shifts (*shifted to*), excluding ongoing shifts (*shifting to*). Language-level controls includes all the baseline controls. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 10: *Male Dominance and Language Survival in Space*

	Dependent variable: language survival (0-1)				
	(1)	(2)	(3)	(4)	(5)
MDI	0.40***	0.40**	0.60***	0.32*	0.39**
	(0.153)	(0.156)	(0.170)	(0.164)	(0.186)
Mean dep. var.	0.490	0.490	0.491	0.487	0.489
Observations	14808	14808	13400	13232	13355
Clusters	741	741	723	712	723
Adjusted R^2	0.016	0.020	0.055	0.078	0.151
Group size		✓	✓	✓	✓
Distance from neighbor		✓	✓	✓	✓
Cell split by country border		✓	✓	✓	✓
Geography and climate			✓	✓	✓
Other controls				✓	
Country fixed effects					✓

Notes: The unit of observation is a grid cell within an ethnic homeland in the Murdock map. "Group size" denotes the logarithm of the number of grid cells forming an ethnic homeland, and varies at the Murdock group level; "distance from neighbour" captures the distance, in degrees, between each grid cell and the closest neighbouring Murdock group; "cell is crossed by country border" is a binary indicator that equals 1 if the cell is crossed by a country border and 0 otherwise; "Geography and climate" are grid-cell level controls including include the CSI (pre 1500), distance from the coast, elevation, ruggedness, climate variability, the malaria index, and a binary indicator that equals 1 if the cell is in tropical climate and 0 otherwise. "Other controls" includes the following cell-level controls: presence of any natural resource, the log of number of natural disasters, the log number of battles; and the following Murdock-level controls: nomadism and an index of political complexity. Robust standard errors clustered at the ethnic homeland level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 11: *Male Dominant Neighbours and Language Survival in Space*

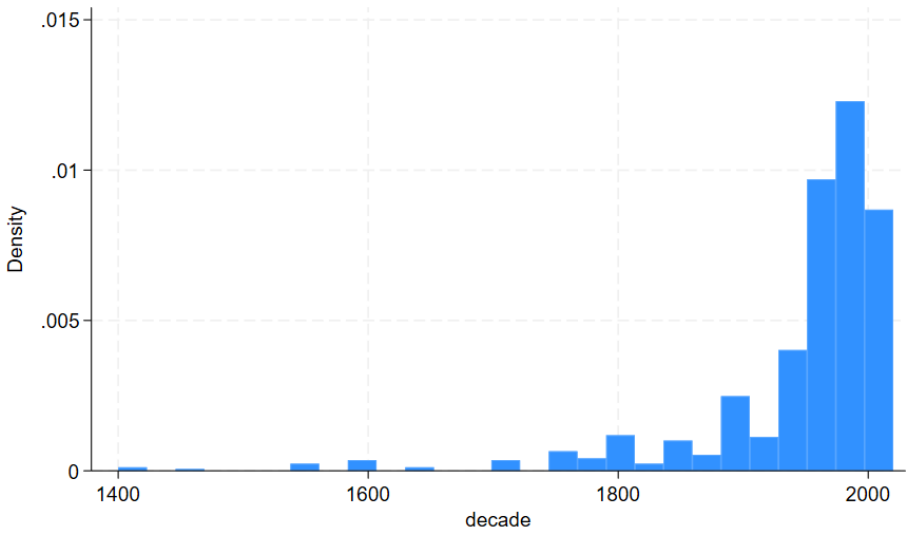
	Dependent variable: language survival (0-1)								
	Average effect			Groups with above-median MDI			Groups with below-median MDI		
	(1)	(2) dist<50km	(3)	(4)	(5) dist<50km	(6)	(7)	(8) dist<50km	(9)
MDI of closest neighbour	-0.15*	-0.15*	-0.19*	0.03	0.07	0.02	-0.37***	-0.37***	-0.30***
	(0.088)	(0.080)	(0.098)	(0.132)	(0.109)	(0.129)	(0.095)	(0.102)	(0.083)
SDI of closest neighbour			0.04			0.04			0.07
			(0.039)			(0.057)			(0.051)
Mean dep. var.	0.479	0.472	0.467	0.519	0.514	0.509	0.433	0.428	0.422
Observations	14699	12295	12879	7822	6322	6396	6877	5973	5357
Clusters	825	824	804	384	384	376	441	440	408
Adjusted R^2	0.549	0.537	0.542	0.579	0.576	0.587	0.518	0.499	0.544
Distance from neighbour	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cell split by country border	✓	✓	✓	✓	✓	✓	✓	✓	✓
Other controls						✓			✓
Homeland fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is a grid cell in a Murdock ethnic homeland. "Distance from neighbour" captures the distance, in degrees, between each grid cell and the closest neighbouring Murdock group; "cell split by country border" is a binary indicator that equals 1 if the cell is crossed by a country border and 0 otherwise; "Other controls" are grid-cell level controls including include the CSI (pre 1500), distance from the coast, elevation, ruggedness, climate variability, the malaria index, a binary indicator that equals 1 if the cell is in tropical climate and 0 otherwise, presence of any natural resource, the log of number of natural disasters, the log number of battles. Robust standard errors clustered at the Murdock ethnic homeland level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Appendix (Intended for Online Publication)

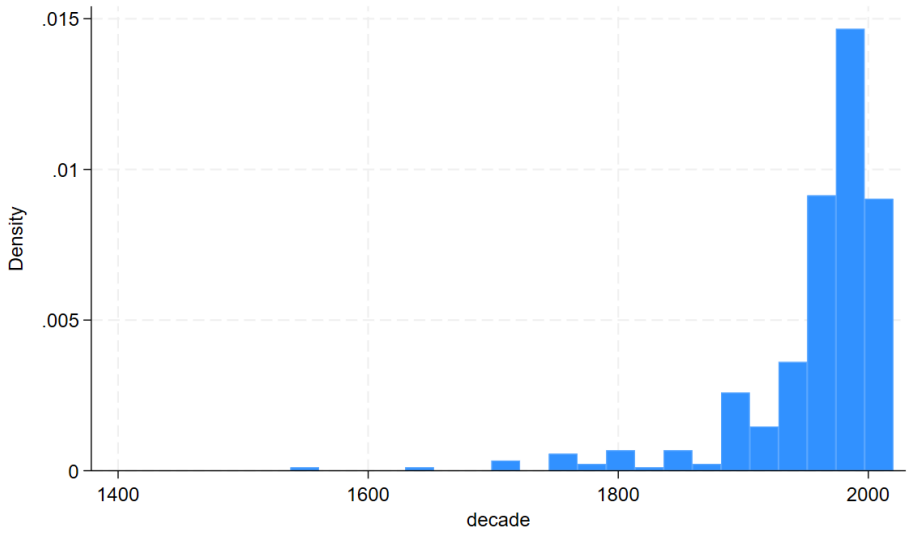
A Additional Figures and Tables

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



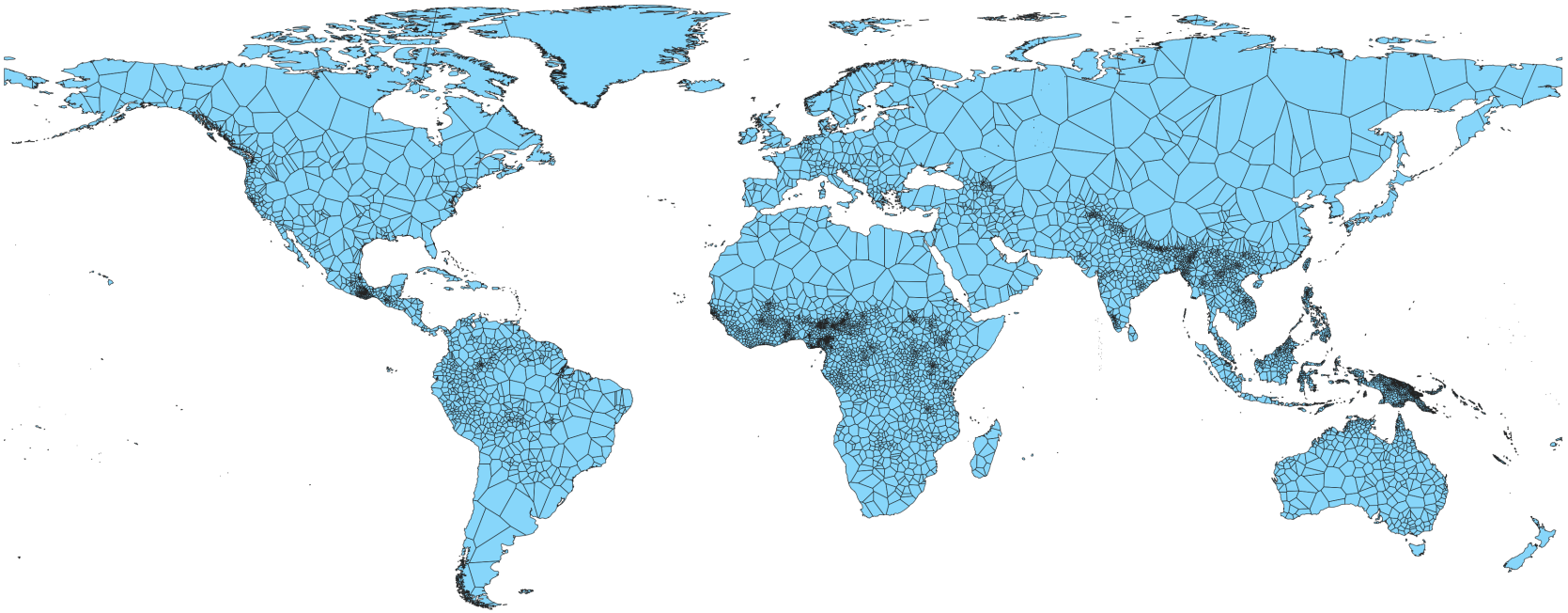
Notes: Based on information on 737 dormant or extinct language-countries (or 722 distinct languages) for which the Ethnologue provides information on the timing of extinction. Dates are approximate and based on when the last known speaker died.

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



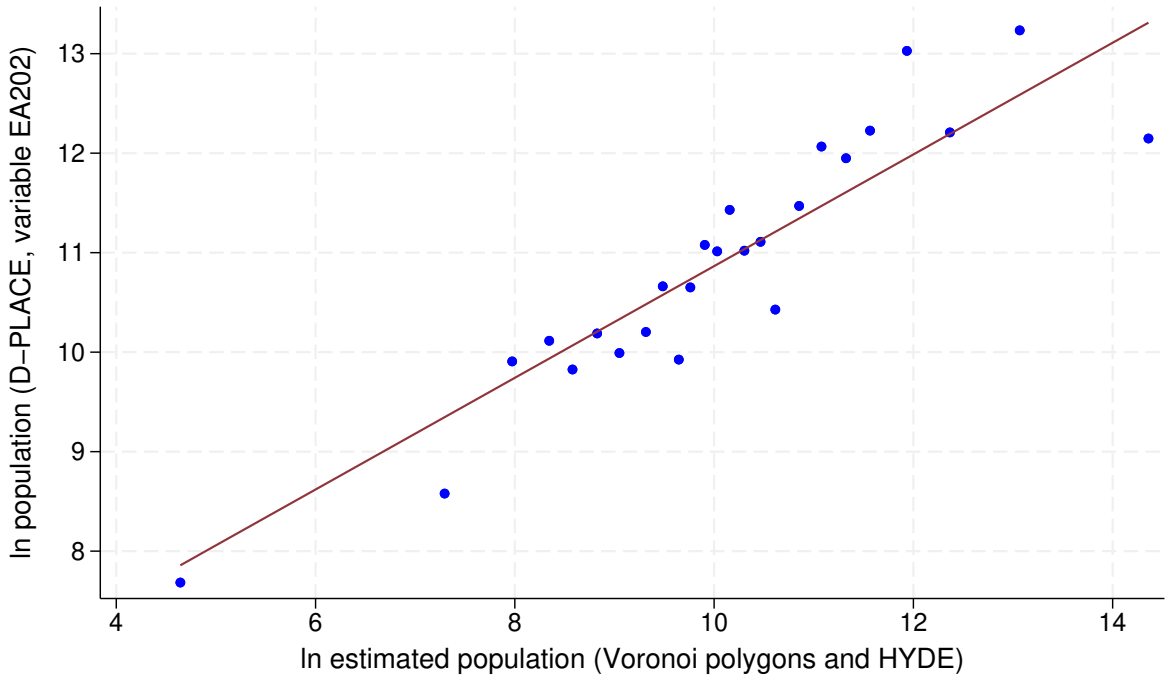
Notes: Based on information on 386 dormant or extinct language-countries (or 383 distinct languages) in our estimation sample for which the Ethnologue provides information on the timing of extinction. Dates are approximate and based on when the last known speaker died.

FIGURE A-3: *Language Voronoi Polygons*



Notes: This map reports Voronoi polygons based on the geographic location (latitude and longitude) of 7,091 living and extinct Indigenous languages. These polygons serve as the geographic unit to estimate Indigenous groups' size at the time of first European colonial contact, by aggregating gridded population data from HYDE to the Voronoi polygon level.

FIGURE A-4: *Estimated Population at First Colonial Contact: Measure Validation*



Notes: The figure shows the binned scatterplot of the natural log of estimated population derived from Voronoi polygons combined with HYDE data (x-axis) against the natural log of population reported in the D-PLACE database (EA202 variable, y-axis) for a sample of 529 Indigenous groups. The correlation between the two variables is 0.53 (p-value < 0.001).

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: *Spatial Analysis Sample: Summary Statistics*

	Mean	St. dev.	Min	Max	N
Language survival	.4839943	.4997592	0	1	16213
Ancestral MDI	.6295148	.1554879	.1835572	.9662781	15891
MDI of closest neighbour (max)	.636403	.1553845	.1835572	.9662781	16311
MDI of closest neighbours (avg)	.6290904	.1547626	.1835572	.9662781	16311
SDI of closest neighbour (max)	.458442	.2436808	0	1	14413
SDI of closest neighbour (avg)	.4423863	.2391749	0	1	14413
Distance from neighbour (in degrees)	.2875072	.3874207	2.60e-07	3.211575	17969
Group size (log # of grids)	2.814518	1.148387	.6931472	5.70711	17969
Cell crossed by country border	.1172018	.3216696	0	1	17969
Absolute latitude	14.95342	9.261887	.25	37.25	17911
Caloric suitability index	819.5884	641.7275	0	2376.317	17963
Distance to the coast	608697.2	456203.9	0	1803629	17969
Elevation	620.038	447.9176	-105.3118	3562	17969
Ruggedness	68.65965	88.48781	0	1271.081	17969
Climate variability	-.0792349	.2875498	-.87	.61	15888
Malaria index	9.566521	10.00689	0	38.08099	17969
Tropical climate (binary)	.4237854	.4941709	0	1	17969
Any natural resource	.109355	.3120928	0	1	17969
Natural disasters (ln)	.1791744	.5123794	0	4.043051	17969
Historical battles (ln)	.0254989	.1745719	0	3.258096	17969
Nomadic	.3428413	.4746736	0	1	16197
Political complexity	1.495883	1.027151	0	4	16030

TABLE A-4: *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
SDI	.3251358	.2526289	0	1	4125
Population size at contact (ln)	8.85447	2.424888	0	15.92227	4289

TABLE A-5: *Horserace: MDI vs Individual 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-6: *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-7: *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-8: *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-9: *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-10: *Language Survival in Space: Measure Validation*

	Dependent variable (from Ethnologue):					
	extinct (0-1)		status (1-12)	extinct (0-1)		status (1-12)
	(1)	(2)	(3)	(4)	(5)	(6)
Language survived in cell	-0.05**	-0.03***	-0.91***			
	(0.025)	(0.011)	(0.161)			
Share of survived cells				-0.03***	-0.03***	-1.16***
				(0.010)	(0.009)	(0.144)
Mean dep. var.	0.025	0.025	4.805	0.010	0.010	4.766
Observations	13643	13643	13531	933	931	913
Clusters	736	736	734	736	735	732
Adjusted R^2	0.027	0.504	0.247	0.012	0.200	0.199
Country fixed effects		✓	✓		✓	✓

Notes: The unit of observation is a grid cell in a Murdock ethnic homeland in columns (1)-(3) and a Murdock group-country in columns (4) to (6). The dependent variable is the extinct indicator based on the Ethnologue database used in the language-level analysis (columns (1), (2), (4), and (5)) or a numerical variable ranging between 1 and 12 constructed based on the EGDIS scale provided by the Ethnologue. In the grid-cell analysis, "language survived in cell" is a language survival indicator that equals 1 if the ancestral language is still spoken in grid cell today and 0 otherwise; in the Murdock-group analysis, "share of surviving in cells" capture the share of ancestral grid cells in which the language is still spoken today. Robust standard errors clustered at the Murdock ethnic homeland level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A-11: *Male Dominant Neighbours and Language Survival in Space: Robustness (I)*

	Dependent variable: language survival (0-1)								
	Average effect			Groups with above-median MDI			Groups with below-median MDI		
	(1)	(2) dist<50km	(3)	(4)	(5) dist<50km	(6)	(7)	(8) dist<50km	(9)
MDI of closest neighbour (average)	-0.12 (0.096)	-0.11 (0.090)	-0.15 (0.107)	0.09 (0.141)	0.13 (0.120)	0.09 (0.139)	-0.38*** (0.102)	-0.40*** (0.114)	-0.30*** (0.089)
SDI of closest neighbour (average)			0.05 (0.044)			0.04 (0.063)			0.09 (0.057)
Mean dep. var.	0.479	0.472	0.467	0.519	0.514	0.509	0.433	0.428	0.422
Observations	14699	12295	12879	7822	6322	6396	6877	5973	5357
Clusters	825	824	804	384	384	376	441	440	408
Adjusted R^2	0.549	0.537	0.542	0.579	0.576	0.587	0.518	0.499	0.544
Distance to neighbour	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cell split by country border	✓	✓	✓	✓	✓	✓	✓	✓	✓
Other controls						✓			✓
Homeland fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is a grid cell in a Murdock ethnic homeland. Robust standard errors clustered at the Murdock ethnic homeland level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A-12: *Male Dominant Neighbours and Language Survival in Space: Robustness (II)*

	Dependent variable: language survival (0-1)								
	Average effect			Groups with above-median MDI			Groups with below-median MDI		
	(1)	(2) dist<50km	(3)	(4)	(5) dist<50km	(6)	(7)	(8) dist<50km	(9)
MDI of closest neighbour	-0.11	-0.16	-0.10	0.14	0.08	0.21	-0.52***	-0.53***	-0.48***
	(0.104)	(0.110)	(0.122)	(0.114)	(0.124)	(0.130)	(0.156)	(0.176)	(0.143)
SDI of closest neighbour			0.04			-0.04			0.10
			(0.049)			(0.065)			(0.064)
Mean dep. var.	0.631	0.621	0.624	0.649	0.638	0.634	0.610	0.603	0.618
Observations	11137	9331	9630	6062	4916	4718	5075	4415	4073
Clusters	770	769	739	347	347	311	423	422	406
Adjusted R^2	0.579	0.555	0.573	0.640	0.614	0.658	0.525	0.506	0.554
Distance to neighbour	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cell split by country border	✓	✓	✓	✓	✓	✓	✓	✓	✓
Other controls						✓			✓
Homeland fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is a grid cell in a Murdock ethnic homeland. The sample is restricted to cells that, in present times, feature at least one language that was originally included in the Murdock Map. Robust standard errors clustered at the Murdock ethnic homeland level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A-13: *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.

TABLE A-14: *Estimated Population at Colonial Contact: Measure Validation*

	Dep. var.: ln population (D-PLACE)		
	(1)	(2)	(3)
ln estimated population	0.56***	0.53***	0.46***
(Voronoi + HYDE)	(0.039)	(0.043)	(0.054)
Observations	529	529	510
Adjusted R^2	0.281	0.322	0.400
Continent FE		✓	
Country FE			✓

Notes: The unit of observation is a Murdock group. The dependent variable is the natural log of EA202 from the D-PLACE database, measuring the total population of the ethnic group. The independent variable is our estimated group population, constructed by overlaying the group's language Voronoi polygon (see Figure A-3) onto HYDE gridded population data for the decade of the D-PLACE population estimate (the date of the estimate is recorded in the variable "year"). Robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

B Data Appendix

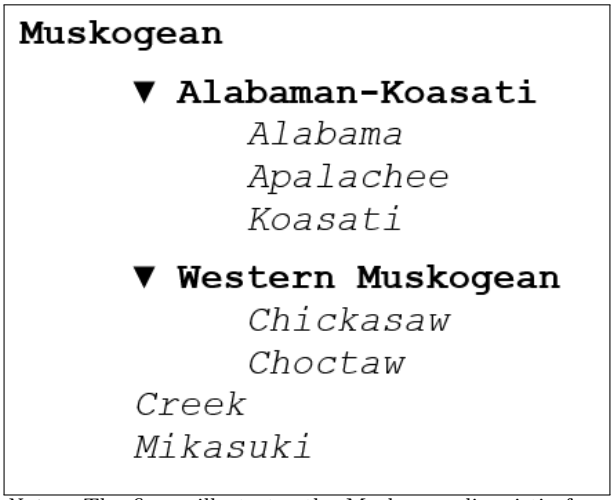
B.1 Procedure to Match Languages to Murdock Ethnic Groups

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-5 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-5: *Example of Linguistic Tree*



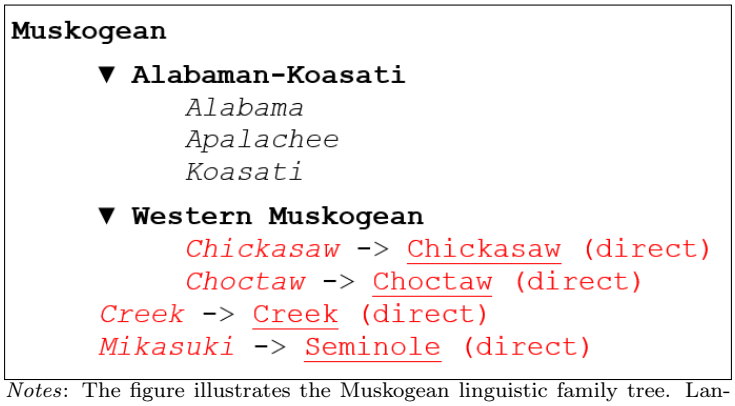
Notes: The figure illustrates the Muskogean linguistic family tree. Languages are the leaf nodes of the linguistic tree, and are indicated in *italics*. Source: `https://glottolog.org/resource/languoid/id/musk1252`

For some of these languages, the D-PLACE database provides us with a direct counterpart in the Murdock EA. Figure A-6 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-6: *Direct Matches between Languages and Murdock Groups*

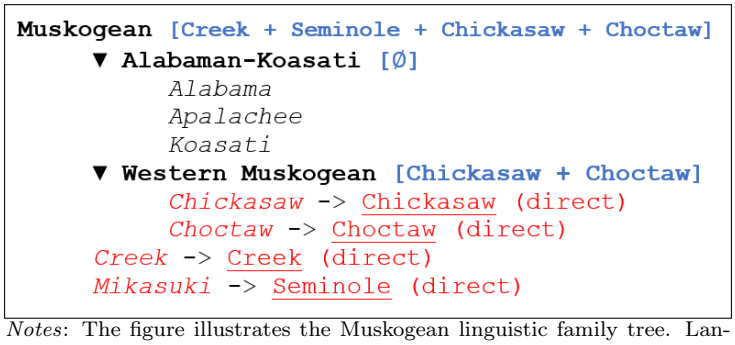


Notes: The figure illustrates the Muskogean linguistic family tree. Languages are the leaf nodes of the linguistic tree, and are indicated in *italics*. Languages that can be directly linked to a Murdock ethnic group are displayed in red. The name of the associated ethnicity is displayed in red and underscored. 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-7 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.

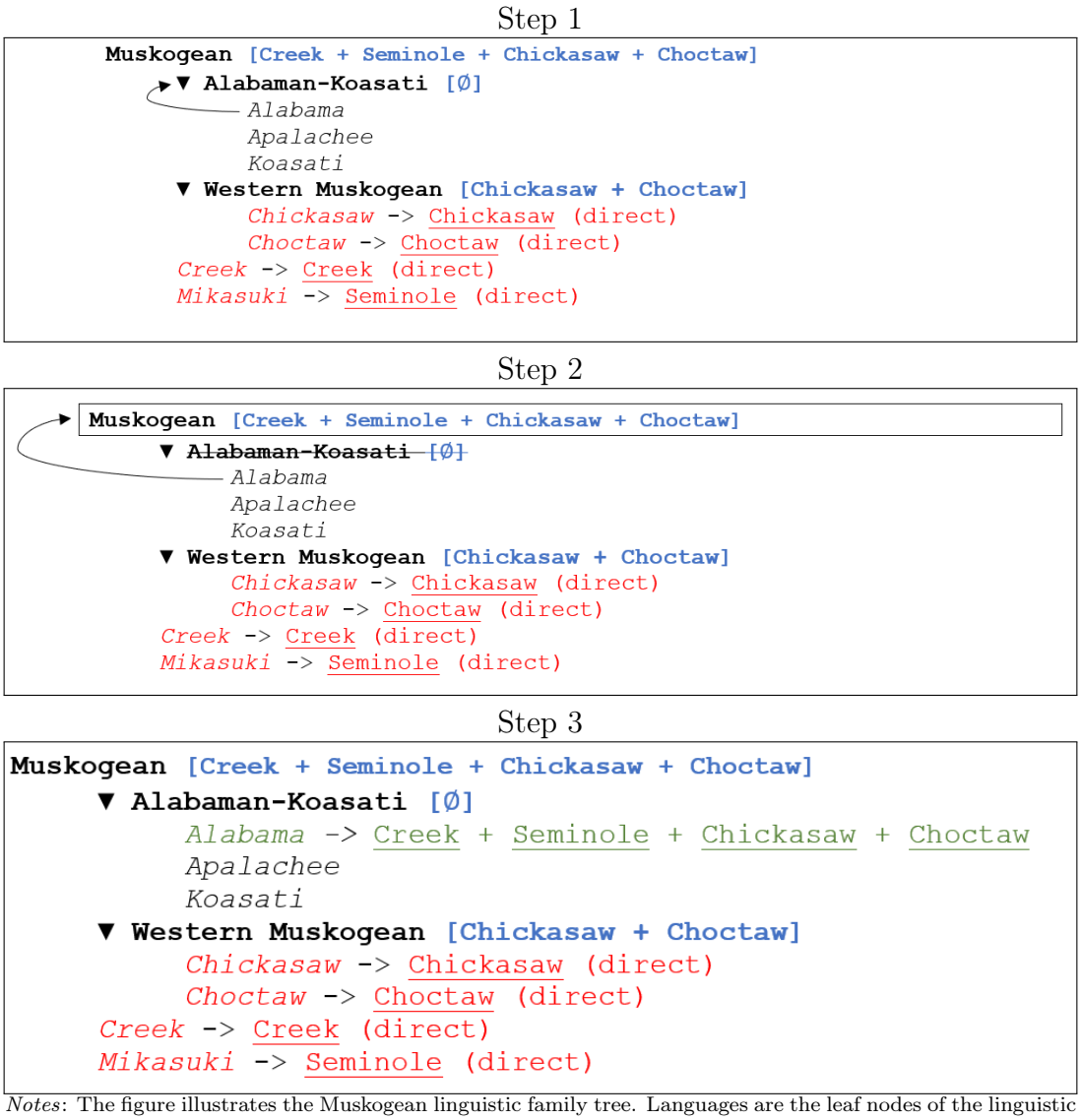
FIGURE A-7: *Assigning Murdock Groups to Nodes in the Linguistic Tree*



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-8 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-8: *Moving up the Tree to Identify Matches*

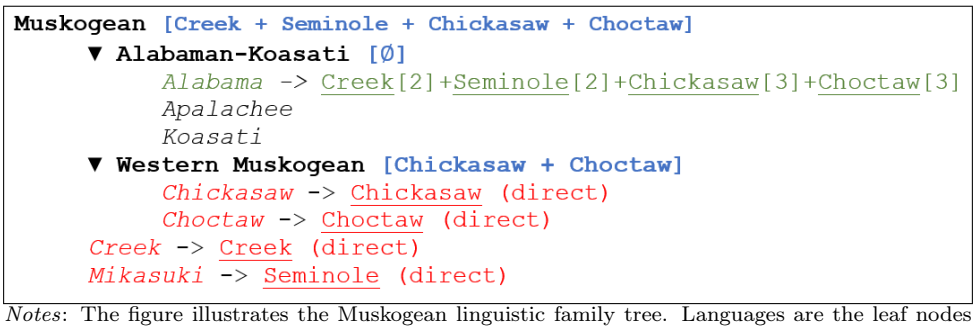


Notes: The figure illustrates the Muskogean linguistic family tree. Languages are the leaf nodes of the linguistic tree, and are indicated in *italics*. Languages that can be directly linked to a Murdock ethnic group are displayed in red. The name of the associated ethnicity is displayed in red and underscored. 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-9 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-9: *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-4 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 combin-

ing the Caloric Suitability Index Developed by Galor and Özak (2016) to the Ethnologue Map and languages' geographic coordinates. We compute the average caloric suitability 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 on 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

Group Size at Colonial Contact: Our estimated measure of group size at colonial contact is based on the HYDE 3.2 database. For a description of the procedure we adopted to construct this measure, please refer to Section 6.1.
Source: https://archaeology.datastations.nl/dataset.xhtml?persistentId=doi:10.17026/DANS-25G-GEZ3&version=2.0