NETWORK LISER

Discussion Paper Series

IZA DP No. 18947

September 2026

Male Dominance and Cultural Extinction

Eleonora Guarnieri
University of Bristol,
CEPR and CESifo

Ana Tur-Prats
University of California, Merced
and IZA@LISER

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18947 September 2026

Male Dominance and Cultural Extinction*

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 ancestral 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 languages, we start by showing that languages from more gender-equal societies face a higher likelihood of extinction than those from male-dominant societies. We propose a conceptual framework in which male-dominant societies are more expansionist via two channels: (i) military capacity sustained through surplus extraction, and (ii) demographic pressure through women's specialization in domestic and reproductive labor. As a result, they expand at the expense of more gender-egalitarian groups. We test this 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. We examine alternative pre-colonial societal arrangements and find that cultural extinction is best explained by asymmetries in male dominance rather than in hierarchy or dominance more broadly. Exploiting local variation in neighboring groups' gender structure across Africa, we further show that languages from gender-egalitarian societies are less likely to survive in grid cells located close to male-dominant neighbors, with military capacity and population pressure predicting language expansion across the continent.

JEL classification

Z10, O10, F54, J16

Keywords

cultural diversity, language extinction, gender norms, inter-group contact

Corresponding author

Ana Tur-Prats

atur-prats@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, Toulouse School of Economics, 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, and 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, 1985; 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.

Gender norms are a central feature of culture, reflecting structural aspects of a society's economic and institutional organization. Male dominance is a societal arrangement in which one group (men) exercises power over another group (women), enabling the extraction of labor and resources (Lerner, 1986). We argue that male-dominant societies are systematically more expansionist than their egalitarian counterparts, owing to two reinforcing features: (i) the extraction of labor and resources can be redirected toward military capacity; (ii) women's specialization in domestic and reproductive work generates demographic pressure.[1] Together, these features generate both the incentive and the capacity to expand.

Whether expansion succeeds depends not only on the expanding group's capacity, but also on the social structure of the target group. Structural differences between competing cultures create asymmetries in expansionist and defensive capacity, and expansion will be more likely whenever male-dominant groups encounter more egalitarian groups. Expansion then drives cultural extinction through biological replacement, the forcible imposition of new norms and institutions, voluntary assimilation, or a combination of these mechanisms. This yields a testable prediction: cultural extinction should be systematically more likely among gender-egalitarian groups exposed to relatively more male-dominant competing cultures. We take this hypothesis to the data across two empirical settings: European colonization and cross-cultural variation across the African continent.

[1]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: male dominance is self-enforcing, less costly to sustain, and closely tied to women's specialization in domestic and reproductive work.

Studying cultural extinction empirically poses a fundamental 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).

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 Scale (EGIDS), with the gender structure of the societies speaking those languages, captured by the Male Dominance Index (MDI) (Guarnieri and Tur-Prats, 2023). The MDI draws on ten pre-industrial 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. 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. The striking empirical patterns shown in Figure 1 suggest that this is the case: in the modern era, language extinction has been most prevalent among groups that were historically more gender equal.[2] When regressing an extinction indicator on the MDI, we confirm this pattern. The negative relationship between male dominance and cultural extinction remains strong and significant after controlling for a

[2]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, nearly extinct, and dormant languages in our estimation sample for which this information is available, 90 percent saw their last speaker die after 1800, with the remaining 10 percent recording their last speaker between 1550 and 1790.

wide set of language-level geo-climatic and historical determinants of economic development (Spolaore and Wacziarg, 2013) and cultural norms (Giuliano, 2026). It also holds when controlling for other bundles of ethnic traits capturing alternative forms of societal dominance and a society's capacity for cooperation. The coefficient survives the inclusion of country fixed effects, which absorb the role of national institutions, colonial history, and national policies. Taken together, these results point to male dominance as a strong and robust correlate of cultural survival, suggesting that gender structure may play a crucial yet overlooked role in the process of cultural evolution.

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 our central hypothesis—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 competing explanations, focusing on group size at colonial contact and distance in other forms of societal dominance. We also rule out a range of additional alternatives, including distance in political complexity, kinship tightness, and linguistic distance.

First, we address the possibility that pre-contact group size might correlate with male dominance and confound our results, as larger populations may have provided a demographic buffer against epidemic mortality and colonial pressure (Thornton, 1987). Since existing ethnographic sources do not provide systematic data on group size at contact, we develop and validate a proxy by combining gridded historical population estimates with the geographic distribution of Indigenous languages, both extant and extinct. 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 one manifestation of a society's hierarchical nature, our results may reflect clashes in societal dominance rather than in gender structure specifically. To test this, we construct a Societal Dominance Index (SDI), defined as the first principal component of pre-colonial political complexity, class stratification, and slavery, capturing the concentration of social and political power in the hands of a few. Distance in the SDI predicts cultural extinction on its own, but its effect attenuates substantially once distance in male dominance is included in the same specification, while the coefficient on the distance on MDI remains stable.

Our results are similarly robust to controlling for three additional dimensions along which Indigenous groups and their colonizers might have differed. Distance in pre-colonial political complexity, a standard proxy for state capacity (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013; Mayshar *et al.*, 2022), predicts cultural extinction on its own but loses its effect when included alongside distance in the MDI. Distance in kinship tightness, which captures how a society organizes cooperation and moral obligations (Enke, 2019), does not robustly predict cultural extinction, and distance in male dominance retains its predictive power when both are included. Distance in the MDI retains its magnitude and statistical significance once controlling for linguistic distance, a widely-used proxy for cultural distance between groups.

Taken together, these results point toward a gender-based explanation for cultural extinction. To explore the underlying mechanisms, we draw on historical narratives and anthropological scholarship documenting how European colonizers imported their male-dominant social and economic structure into the territories they colonized. This took concrete forms. For instance, among the Cherokee, colonial reformers required men to replace women as farmers, outlawed polygamy, undermined matrilineal kinship, and imposed the nuclear family (Perdue, 1998); in colonial Nigeria, the British excluded Yoruba women from the state structures they had previously occupied (Oyěwùmí, 1997). Similar dynamics are documented across colonial encounters in North and South America (Devens, 1986; Holly, 1990; Buenaventura-Posso and Brown, 1980; Jaimes and Halsey, 1992). 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.

Extinction may reflect the assimilation of a community into the colonizer's culture, the demographic collapse of its speakers, or both. While we cannot fully separate these channels, we leverage Ethnologue data on language shift to provide evidence on the first. We code whether a language is shifting or has shifted to the colonizer's language or to a different language. This approach is informative about cases where language communities abandon their ancestral language while remaining present as a population. 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 and not other languages, suggesting that male-dominance asymmetries shaped susceptibility to colonial assimilation rather than general patterns of language change.

We further test our hypothesis in the context of inter-group contact at the local level in Africa. The availability of ethnolinguistic maps at two points in time for the continent—the pre-colonial period (Murdock map) and the present day (Ethnologue map)—enables the construction of an indicator at the grid-cell level capturing whether the ancestral language has survived up to the present. The continent's exceptional cultural diversity gives rise to within-group, cross-cell variation in the gender structure of neighboring groups. The fine spatial granularity of the data allows us to include homeland fixed effects, so that identification comes from comparing grid cells within the same ancestral homeland. These fixed effects net out group-specific factors, such as absolute defensive capacity of a group or its internal micro-level processes of cultural transmission. We find that the likelihood of language survival tends to be lower in cells neighboring male-dominant groups than in those neighboring gender-equal groups on average. Importantly, the 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.

Finally, we demonstrate that male-dominant groups are systematically more expansionist than gender-equal groups and uncover the mechanisms through which male dominance translates into language expansion. We focus on two channels: the accumulation of military capacity, facilitated by the extraction of surplus, and population pressure arising from women's specialization in domestic and reproductive labor. To test these mechanisms, we construct a novel measure of language expansion at the ethnolinguistic group level. By combining the historical homeland boundaries from the Murdock map with contemporary language distributions from the Ethnologue map, we measure the extent to which groups have expanded beyond their ancestral territories since the pre-colonial period. Three complementary sets of results support our hypothesis. First, male dominance is highly correlated with expansion, more so than the index of social dominance. Second, male-dominant groups display higher levels of military capacity as well as fertility, both measured using folklore-based indicators. By contrast, social dominance is strongly associated with military capacity but not with fertility, consistent with the idea that demographic pressure arises specifically from the gendered division of labor rather than from hierarchical organization more generally. Third, both military capacity and fertility independently predict expansion, suggesting that male-dominant groups expand both because of their military capacity and because they face stronger demographic incentives to acquire new territory.

This paper contributes to various strands of literature. First, by proposing and empirically testing a novel hypothesis on the determinants of language extinction, we contribute 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 analysis of 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 gender-structure mismatches with a competing group place more gender-equal societies at heightened risk of linguistic extinction.

Second, this paper contributes to a long-standing literature on historical gender relations and their far-reaching consequences for social and economic life. A foundational tradition in anthropology and history has documented that gender hierarchies vary substantially across societies (Sanday, 1981; Chafetz, 1984) and that patterns of male dominance, rather than being universal, have emerged from historical processes and colonization (Lerner, 1986; Leacock, 1981). In economics, a growing literature has documented the historical origins of gender norms and their persistent consequences for contemporary outcomes (Alesina *et al.*, 2013; Becker, 2022; Giuliano, 2018; Matavelli *et al.*, 2026; Anderson, 2025). Our paper extends this literature in a new direction. We show that pre-colonial gender structures determine which cultures survive or go extinct: male dominance matters not only within societies, but also for the outcomes of contact and competition between them.

Third, the focus on the between-society dimension connects our paper to the literature on the consequences of inter-group contact. While economists have documented the long-run economic, institutional, and cultural consequences of such contact (Acemoglu *et al.*, 2001; Dell, 2010; Nunn and Wantchekon, 2011; Jha, 2013), its role in determining which cultures survive has received little systematic attention. We introduce cultural survival—measured by whether a group's language persists—as an outcome of inter-group encounters, and document a specific asymmetric pattern: male-dominant groups tend to expand at the expense of gender-egalitarian ones. We show this pattern across two distinct forms of contact. In the context of European colonization, existing work has examined how differences between colonizers and Indigenous groups affect the latter's long-term economic performance (Cornell and Kalt, 1995, 2000; D'Amato and Russo, 2026). We instead analyze how cultural and institutional differences shape their survival and show that the male-dominance gap is the strongest predictor of cultural extinction. Our African spatial analysis is closest in spirit to McGuirk and Nunn's (2025) work on inter-group conflict between pastoralist and agricultural societies, but shifts the focus from conflict to language survival, and identifies gender structure as a key driver of the outcomes we observe.

Finally, our paper speaks to a growing body of work on cultural transmission and evolution.[3] Research in economics has built on models of social learning and cultural socialization to study how norms and values emerge, persist, and change (Bisin and Verdier, 2001, 2026). 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; 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). This literature, however, has focused primarily on the emergence, persistence, and transmission of cultural traits, rather than on the disappearance of the cultural groups that carry them. The evolutionary anthropology tradition, by contrast, has long argued that cul-

[3]See Lowes (2024) and Fernández (2025) for recent reviews of this literature.

tural evolution cannot be understood without accounting for cultural loss and inter-group competition (Boyd and Richerson, 1985; Henrich, 2004, 2016). Cultural group selection theory, in particular, models how distinct groups fare when they compete, but has done so primarily through theoretical and case-study work. We provide the first global empirical analysis of the disappearance of established cultural entities and the inter-group mechanisms that drive it.

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 and extractive 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).

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 defense (Spencer, 2010; Turchin *et al.*, 2013).[4] Whether expansion succeeds depends not only on the expanding group's capacity, but also on its interaction with the target group. Structural differences in societal dominance 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 comparative advantage in territorial expansion, making it more likely whenever they encounter more egalitarian groups (Tilly, 1992).[5]

[4] Hierarchical structures typically involve the extraction of surplus through taxation, tribute, or coerced labor. The degree to which surplus is militarized rather than redistributed may vary across societies, but the link between hierarchy and extractive capacity implies that more hierarchical societies will tend, on average, to have greater potential for expansion.

[5] 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).

2.2 Male Dominance versus 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.

First, *male dominance is self-enforcing 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 might therefore require stronger coercive infrastructure and substantial institutional capacity, 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 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, institutions, 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. The latter 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-described 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.[6]

A more recent illustration, closer to our empirical setting, is the role of gender structure in European colonization. Anthropological and historical 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.

[6] 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).

Buenaventura-Posso and Brown (1980) describe this process in the case of the Barí of Colombia, and Appendix C presents a collection of historical narratives on the colonial transformation of Indigenous gender arrangements. Such disruptions likely eroded the social fabric of gender-egalitarian Indigenous groups, contributing to the process of cultural extinction, that 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, social and cultural norms, geo-climatic characteristics and historical arrangements.

3.1 Languages and their Status

We begin from the ISO 639-3 database curated by SIL International, which lists 8,552 distinct human languages, both living and extinct, and identifies each with a unique three-letter code. ISO 639-3 defines language boundaries using both linguistic criteria (e.g., intelligibility) and considerations of shared ethnolinguistic identity among speakers.[7]

To measure language vitality, we draw on the Ethnologue database (Eberhard *et al.*, 2024), which builds on the ISO 639-3 classification and covers 7,678 of the 8,552 languages.[8] For each language spoken in a given country, the Ethnologue assigns one of 12

[7]Detailed information on the ISO 639-3 classification, including the treatment of dialects and the exclusion of avoidance, secret, ceremonial, play, urban, youth, and special-purpose languages, is provided in Appendix B.2.

[8]The Ethnologue excludes ancient, constructed, and historical languages.

values of the Expanded Graded Intergenerational Disruption Scale (EGIDS), where higher values indicate greater disruption to intergenerational transmission (Lewis and Simons, 2017). Table A-1 reports the full scale: lower values correspond to national languages (e.g., English in the United States, Khmer in Cambodia) or languages in vigorous use (e.g., Basque in Spain, Krio in Sierra Leone), while higher values indicate increasing endangerment. We code a language in a given country as *extinct* if it falls into one of three categories at the top of the scale: *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, as with Dakota in Canada), *dormant* (the language serves as a heritage identity marker for an ethnic community but lacks proficient speakers, as with Karuwali in Australia), or fully *extinct* (no longer used, with no speakers retaining any associated ethnic identity, as with Puquina in Peru). Figure 2 shows the distribution of languages and average MDI by status. The latter reinforces the pattern in Figure 1: less endangered languages tend to display higher levels of male dominance.

3.2 Male Dominance and Other Ethnic Characteristics

Male Dominance. Our primary measure of gender structure is the Male Dominance Index (MDI) developed by Guarnieri and Tur-Prats (2023), constructed from variables in the Ethnographic Atlas (EA), which describes traditional cultural practices for more than 1,200 societies measured prior to industrialization. Informed by interdisciplinary literature on the origins of societal gender norms, the MDI combines ten EA traits related to family arrangements, subsistence activities, and inheritance patterns: matrilineality, polygyny, dependence on fishing, gathering, non-herding animal husbandry, shifting agriculture, bride price, plough use, nuclear family structure, and dependence on pastoralism. The index is the first principal component of these ten variables, rescaled to the unit interval, and loads positively on traits traditionally associated with male dominance and negatively on those linked to greater gender equality. Guarnieri and Tur-Prats (2023) validate the MDI against contemporary survey-based measures of gender (in)equality including female employment, son preferences, gender attitudes, and domestic violence.

Other Indices of Ancestral Social Structure. Alongside the MDI, we construct two additional indices using the same approach—extracting the first principal component of a bundle of EA traits and rescaling to the unit interval—each capturing a distinct dimension of pre-colonial society. First, a Societal Dominance Index (SDI), built from political complexity, class stratification, and slavery. While political complexity captures the hierarchical organization of a polity and its state capacity, class stratification and slavery add the dimension of institutionalized social inequality and the subordination of certain groups to others (Michalopoulos, 2012; Lagerlöf, 2009). Together, these three variables capture not only the organizational sophistication of a society but also the degree to which power is concentrated in the hands of a dominant group.[9] Second, the Kinship Tightness Index (KTI) by Enke (2019), built from four binary variables: the presence of extended (versus nuclear) family households, postmarital residence with the husband's or wife's group, unilineal (versus bilateral) descent, and localized clans. Although these variables describe family and community organization, Enke (2019) shows that they index a broader, internally consistent moral system and summarizes the way a society organizes cooperation.

Other Pre-Colonial Characteristics. From the EA we also construct an indicator of settlement practices (nomadic vs. sedentary) and a standalone measure of pre-colonial political complexity (ranging from 0 to 4, from no political authority beyond the community to large states).

Matching Languages to Ethnic Groups. To assign EA-based measures to languages, we extend the procedure of Bahrami-Rad *et al.* (2021) to extinct languages. Using the D-PLACE database, which maps 1,209 languages to 1,291 EA societies, we create direct links between languages and EA groups; directly linked languages inherit their group's values. Languages without a direct link are assigned the values of their nearest direct link(s) in the linguistic tree. Appendix B.1 describes this procedure in detail. Data on the identity of the colonizer comes from the Colonial Dates (COLDAT) Dataset (Becker, 2019), provided at the country level, and we assign to each colonizer the MDI and other ethnic characteristics of its corresponding EA group. Whenever a linguistic group was colonized by multiple colonial powers—which is the case for 17% of languages in our estimation sample—we take the average MDI of all the colonizers of the language's country.

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 first principal component loads positively on all three variables and explains 62% of the total variance. Loadings: 0.6335 (political complexity), 0.6408 (class stratification), 0.4336 (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] Appendix B.2 describes all data sources, matching procedures, and construction details.

Our covariates fall into four groups. *Geography and caloric suitability* includes absolute latitude, average elevation and ruggedness (U.S. Geological Survey, 1996), distance to the coast (Natural Earth, 2010), and the caloric suitability index of Galor and Özak (2016). *Climatic and disease environment* includes temperature variability for the period 500-1900 (Mann *et al.*, 2009), the share of a language's area in a tropical climate (Beck *et al.*, 2018), and malaria suitability (Kiszewski *et al.*, 2004). *Natural resources and disasters* includes a binary indicator for the presence of gold, petroleum, diamonds, or gemstones (PRIO-GRID), and exposure to natural disasters since 1900 (Centre for Research on the Epidemiology of Disasters, 2023). *Historical conflict* measures the number of battles a linguistic group was exposed to between 1468 BCE and the end of the 18th century, using the Historical Conflict Event Dataset (Miller, 2022).

3.4 Historical Population

Comprehensive data on the size of Indigenous groups at contact is not available at a global scale. We therefore develop a proxy that combines 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 (Figure A-1). 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, dated at the country level by COLDAT.

We validate our Voronoi–HYDE procedure against variable EA202 in D-PLACE, which records ethnic group population for a subset of EA groups.[11] The two measures are

[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.

[11] The D-PLACE population year does not necessarily coincide with the society's year of observation or its pre-colonial period; we construct our Voronoi–HYDE estimate using HYDE data for the same decade as the D-PLACE variable.

positively correlated across the 683 group-countries for which both are available, and the relationship is robust to continent and country fixed effects (Figure A-2 and Table A-2). We are therefore confident in using the Voronoi–HYDE measure as a proxy for Indigenous group size at first colonial contact.

3.5 Summary Statistics

The unit of observation in the resulting dataset is a language-country. Our estimation sample comprises 4,794 languages spoken in 212 countries (N=5,891 observations). Table A-3 in the Appendix reports summary statistics. 11 percent of the language-country observations in our sample are extinct according to our definition. Among these 626 observations (corresponding to 604 languages), 229 (or 36.6%) are classified as extinct, 217 (or 34.7%) as nearly extinct, and 180 (or 28.8%) as dormant.

For a subset of dormant and extinct languages, the Ethnologue provides information on the timing of extinction, approximated by the death of the last known speaker.[12] Figures A-3 and A-4 in the Appendix show the distribution of extinctions over time, both for all languages in the Ethnologue and for those in our estimation sample, respectively. The endpoint of decline of most languages occurred within the last two centuries.

4 Male Dominance and Cultural Extinction

4.1 Empirical Strategy

Our first goal is to establish whether a society's degree of male dominance is an important correlate of its likelihood of cultural extinction. 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] The vector $\mathbf{X}_{ic}$ comprises language-by-country controls organized into six groups, encompassing a comprehensive set of factors that could potentially confound the analysis.

[12] Information on the timing of extinction is available for 384 out of the 604 languages that we classify as extinct in our estimation sample.

[13] The MDI is specific to a language, not a language-country, since the Murdock Ethnographic Atlas underlying it codes ethnic groups irrespective of whether they span multiple countries.

These controls are motivated by existing literature linking them to economic development (Spolaore and Wacziarg, 2013), the formation of cultural norms (Giuliano, 2026), ethnolinguistic diversity, and societal collapse.

Geography: absolute latitude, distance to the coast, ruggedness, elevation, and the caloric suitability index. Following Gallup *et al.* (1999) and Hall and Jones (1999), who document a correlation between distance to the Equator and economic performance, absolute latitude is commonly employed in comparative development literature to account for geographic effects (Acemoglu *et al.*, 2001; Michalopoulos, 2012; Rodrik *et al.*, 2004). Distance to the coast and ruggedness proxy for trade access and isolation, factors that may underlie economic development (Nunn and Puga, 2012; Gallup *et al.*, 1999) as well as language extinction directly. Elevation and the caloric suitability index account for confounders related to agricultural productivity, a fundamental determinant of economic development (Galor and Özak, 2016) and ethnic diversity (Michalopoulos, 2012).

Climate and disease: climate variability, tropical climate, and the malaria index. Nettle (1998) shows that climate variability is a key factor explaining the global distribution of linguistic diversity, through its effect on the seasonality of food production. Tropical climate and disease burden, in turn, are linked to differential income levels and growth through their impact on productivity and institutions (Gallup *et al.*, 1999; Acemoglu *et al.*, 2001).

Natural resources: an indicator for the presence of gold, petroleum, diamonds, or gems in a language's polygon. The presence of natural resources has been shown to affect institutions, economic growth, and violence (Humphreys, 2005; Ploeg, 2011), factors that may matter for cultural extinction as well.

Natural disasters: the log of one plus the number of natural disasters in a language's polygon. Environmental stress and natural disasters have been identified as contributing factors to societal vulnerability and collapse (Diamond, 2011; Butzer, 2012).

Exposure to war: the log of one plus the number of battles 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]

Pre-colonial institutions: ancestral political complexity and settlement patterns (seden-

[14]As discussed in our conceptual framework, conflict is also part of 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.

tary vs. nomadic), to disentangle the role of male dominance from that of institutions (Alesina and Giuliano, 2015; Bisin and Verdier, 2017). Such institutions may define societies' development trajectories (Michalopoulos and Papaioannou, 2013).[15]

We include these sets of controls sequentially, from the arguably most exogenous and predetermined (e.g., geography) to the more endogenous ones, i.e., those that are likely co-evolving with gender 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 Main Results

Table 1 reports our estimates of λ. We find a strong and significant relationship between gender structure and the likelihood of extinction, confirming the visual pattern in Figure 1: the more male dominant a society's deep-rooted gender structure, the lower its language's likelihood of extinction. Column 1 shows that moving from the most gender-egalitarian to the most male-dominant society decreases the probability of extinction by 47 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, gender-egalitarian societies may have systematically emerged in geographies less conducive to long-run development, in regions more exposed to geo-climatic shocks, or in areas lacking growth-inducing natural resources. If so, 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, columns 2–5 show that the MDI remains a strong and stable predictor of extinction after netting out a broad set of geographic and geo-climatic confounders.

Similarly, gender structure may co-evolve with institutional or organizational features of a society that could 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

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

one standard deviation increase in the MDI decreases the likelihood of cultural extinction by 0.22 standard deviations.

4.3 Within-Country Variation

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 0.10 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.

4.4 Interpreting the MDI as a Measure of Gender Structure

The interpretation of our results depends on the MDI capturing a society's gender structure rather than other dimensions of pre-colonial society. We propose three exercises to support our interpretation. First, the MDI bundles ten ethnic traits into a single index, which raises the possibility that the relationship with extinction could be driven by one specific trait rather than by gender structure as a whole. We address this by running horse-races between the MDI and each of its ten constituent traits, one at a time. Table A-6 shows that, across all ten specifications, the MDI coefficient remains negative and statistically significant, suggesting that the MDI captures variation that no individual component can account for on its own.

Second, the MDI's predictive power could reflect properties of this particular combination of EA variables, raising the possibility that we capture something about ancestral economic and family organization rather than gender structure as such. We further validate the MDI against an entirely different measure of gender norms: the Male Dominance Bias (MDB), developed by Michalopoulos and Xue (2021) from ethnic-specific oral tradition describing men as more dominant, less submissive, less engaged in domestic affairs, and more physically active relative to women.[16] The MDB and the MDI are strongly positively correlated, both unconditionally and conditional on the full set of controls (Table A-7), suggesting that the two indices, despite being constructed from different data sources, capture a common underlying dimension of gender structure. Table A-8 further

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

shows that replacing the MDI with the MDB in our main specification produces qualitatively similar, albeit less precisely estimated, results: more male-dominant oral traditions are associated with a lower likelihood of cultural extinction.

Third, while the MDI captures gender structure as a dimension of social organization, it might correlate with other social and cultural dimensions that could drive cultural extinction. For example, the MDI may correlate with other kinship and marriage traits not included in the index. Two natural candidates are patrilocality and exogamy. When spouses come from different language groups, both practices increase the linguistic heterogeneity of households and could thereby accelerate intergenerational language shift in favor of the established kin group at the expense of the outsider's language. Table A-9 controls for patrilocality and exogamy alongside the MDI. The MDI coefficient is essentially unchanged whether either or both variables are included, and neither predicts extinction independently.

We also consider two other key dimensions of societal organization: the concentration of political and social power, measured through the Societal Dominance Index (SDI), and the organization of cooperation and morality, captured by the Kinship Tightness Index (KTI) (Enke, 2019). Both indices are constructed analogously to the MDI, as the first principal component of different bundles of EA traits. Table A-10 reports the cross-correlations among the three indices.[17] Table A-11 in the Appendix shows that the MDI's predictive power is robust to the inclusion of both alternative indices. Entered individually, the MDI and SDI both predict extinction, while the KTI does not. In a horse-race specification, the MDI coefficient remains large and significant, the SDI coefficient attenuates substantially, and the KTI coefficient is small and insignificant once language-level controls are included.[18]

[17] In contrast to societal dominance, which is strongly positively correlated with male dominance ($\rho = 0.67$), kinship tightness is negatively correlated with both the MDI ($\rho = -0.38$) and the SDI ($\rho = -0.24$). This indicates that the KTI captures variation in social organization that is largely orthogonal to the gender-norm and political-power dimensions at the heart of our argument.

[18] For this exercise, we include all language-level controls with the exception of pre-colonial institutions (nomadism and political complexity). Because tight or loose kinship regimes might confer survival advantages through different mechanisms operating at different scales, a flat linear coefficient could in principle mask a non-linear relationship. We therefore test for non-linearity by adding a quadratic term for KTI in Table A-12. The estimated relationship is not robust. Without controls, the KTI enters with a positive linear and negative quadratic term, implying an inverted-U, a pattern that disappears or reverses sign as soon as language-level controls, the MDI and the SDI are added. Crucially, the MDI coefficient is stable across all specifications, and remains unchanged whether the KTI is omitted, linear, or quadratic.

4.5 Matching Procedure and the Issue of Selection into the Sample

As explained in Section 3, the MDI is assigned to languages through proximity in the linguistic tree. By construction, languages with no relatives directly matched to the EA are less likely to receive an MDI value and thus drop out of our estimation sample. Language isolates, i.e., languages that do not belong to any linguistic family, are a case in point as they have a lower chance to have a directly-matched relative: only 31% of language isolates receive an MDI, while 76% of non-isolates do. Moreover, extinction rates differ across matched and non-matched languages: among languages that our tree-based algorithm is unable to match, the extinction rate is 5.4 percentage points higher than for languages in our estimation sample. Since we do not know where these languages sit on the MDI distribution, their absence could in principle bias our estimates in either direction. We address this selection issue in two ways.

First, because extinction is observed for every language in the Ethnologue irrespective of whether we are able to match them to the EA or not, we can reintroduce 2,353 unmatched languages (or 2,490 language-countries) under any assumed MDI value and trace the estimated coefficient across the full support of the index. Figure A-5 reports the result of this exercise. Each point is a separate estimation of equation 1 in which every unmatched language is assigned the value of the MDI displayed on the horizontal axis. The coefficient is negative at every value in the MDI support and significant at the 5% level for any assigned MDI value below 0.89. This threshold of 0.89 lies at the 99th percentile of the MDI distribution among languages in our estimation sample. Therefore, overturning our result would require assuming that the languages we cannot match are more male-dominant than 99 percent of the languages we are able to match.

Second, we assign to the unmatched languages the MDI value of their geographically nearest neighbor.[19] This methodology might introduce some degree of measurement error into the MDI for these languages, especially if the closest language group is located several hundred kilometers away and may thus belong to an entirely different cultural setting.[20] We nevertheless re-estimate our baseline specification using this broader sample of languages and report estimates in Table A-13. Column 1 displays our baseline finding using the tree-based MDI matching for comparison. Column 2 shows that adding

[19]We do the same for political complexity and nomadism, i.e., the two controls that come from the EA and thus rely on our matching procedure.

[20]The average distance between an unmatched language and its closest neighbor is 149 km (median 97 km), with a maximum of 2,085 km.

initially unmatched languages through the nearest neighbor approach leaves our conclusions unchanged: albeit somewhat smaller, the relationship between male dominance and extinction remains large and significant. Similar findings arise when restricting the imputation to languages whose nearest neighbor lies within 100 km (column 3).

Finally, we discard our tree-based matching algorithm altogether, and test whether our results hold under a nearest neighbor matching algorithm. Using the "direct links" from D-PLACE, we assign to each language the MDI value of its geographically closest language.[21] This approach rests on a different assumption about how culture diffuses: while our baseline measure assumes, as is common in the literature, that culture travels with linguistic descent, the geographic measure assumes that it is shared across neighboring populations regardless of ancestry. Results with this alternative matching are reported in columns 4 and 5 of Table A-13. The geographic measure yields a similar coefficient to our baseline measure, and focusing only on languages with neighbors within 100 km leaves it unchanged. We therefore find no evidence that our results depend on the assumption underlying the matching procedure.

4.6 Additional Robustness

Our results are robust to a battery of additional checks. Table A-14 shows that the MDI coefficient remains negative and significant when restricting the sample to languages indigenous to a country (columns 1–2), when we employ an alternative matching procedure between the MDI and languages (columns 3–4),[22] when we restrict the sample to languages that can be directly matched to a Murdock ethnicity via D-PLACE (columns 5–6), or when we cluster standard errors by language family rather than country (columns 7–8). 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 9 and 10).

Table A-15 shows that results also hold when employing wider or narrower measures of extinction as dependent variables, i.e., one that includes shifting and moribund languages (columns 1–2) and one that only considers fully extinct languages (columns 3–4), respectively. Columns 5 and 6 remove from the sample languages that have gone extinct before

[21]If no MDI information exists for that closest language, the resulting MDI is set to missing, as in our tree-based algorithm.

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

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.

The patterns documented in this section establish a strong and robust correlation between deep-rooted gender structure and cultural extinction. The MDI's predictive power survives an extensive set of controls, holds within country, is not driven by any single component of the index, replicates with a folklore-based measure of male dominance, and is not displaced by alternative dimensions of ancestral social organization. The breadth of the robustness suggests that male dominance is a meaningful and distinctive predictor of which cultures survive. The next sections analyze inter-group dynamics by leveraging two empirical settings: European colonization and variation in neighboring groups' gender norms across Africa. This allows us to move closer to identification and provide more direct evidence on the underlying mechanisms.

5 Distance in Male Dominance, Colonization, and Cultural Extinction

European colonization provides a unique setting to test the asymmetries in capacity argument at the heart of our conceptual framework. It exposed the majority of non-European cultural groups to a new competing dominant culture. Linguists describe colonization as one of the major causes of cultural extinction in recent times (Mufwene, 2002), and Lugones (2016) argues that gender relations in pre-colonial societies are key to understanding the scope of the social transformation enacted by colonization.[23]

5.1 Empirical Strategy

We relate each Indigenous language to its colonizer in a dyadic setting and consider the role of cultural distance in male dominance in explaining extinction. We restrict the sample to 3,862 Indigenous languages in 106 countries (N=4,339) affected by European colonization.[24] Table A-4 presents summary statistics for this sample. On average, the

[23]"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 heterogeneous 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.

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

prevalence of extinction is slightly higher compared to that in the full sample of languages described in Table A-3, while the level of male dominance is lower.

For each Indigenous language i in country c, we compute the absolute distance in male dominance between the language and its colonizer:

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

where MDI_c is the colonizer's MDI.[25] We estimate:

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

where $\mathbf{X}_{ic}$ is the full set of language-level controls discussed in Section 4. This specification compares the explanatory power of the Indigenous language's *absolute* level of male dominance with that of its *distance* from the colonizer.

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

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

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

The decomposition allows us to test the asymmetric-capacity prediction from our conceptual framework: the two coefficients should differ, with extinction risk concentrated among Indigenous groups that were less male-dominant than their colonizers. We estimate:

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

We also re-estimate this equation controlling for the colonizer's MDI (MDI_c) rather than the language's own.

A natural extension of this dyadic design is to leverage variation across Indigenous languages exposed to the *same* colonizer. For instance, while the Native American Keres

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

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. We exploit this variation by including colonizer fixed effects (τ_c) in our regression:

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

The fixed effects (τ_c) absorb all observable and unobservable characteristics of the colonizing power—its gender structure, cultural traditions, political and economic institutions, colonial policies, and military power—thereby tightening identification. The asymmetry is identified from variation among Indigenous groups exposed to the same colonizer. Standard errors are clustered at the country level throughout.

5.2 Results

Tables 3 and 4 present the results. Column 1 of Table 3 shows that the strong and significant relationship between a language's level of male dominance and its likelihood of extinction persists in the sample of Indigenous languages affected by European colonization. Once we include the absolute distance in gender structure from the colonizer (column 2), however, the coefficient on the language's own MDI drops in magnitude and loses statistical significance, while the coefficient on distance is positive and significant: the larger the distance in male dominance from the colonizer, the higher an Indigenous language's risk of extinction. This suggests that cultural distance is a stronger predictor of extinction than the language's absolute level of 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 asymmetric pattern emerges: a language's risk of extinction increases only whenever its gender structure is more egalitarian than that of the colonizer, but not the reverse. 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 and statistically significant at the 5 percent level. We caution against over-interpreting the small and insignificant coefficient on *Language More MDI* as evidence of a null effect in that direction. Colonizers were on average substantially more male-dominant than the Indigenous groups they encountered, so there is little variation in this variable in our sample.

Table 4 shows that the same pattern holds when controlling for the colonizer's MDI or including colonizer fixed effects. Column 1 reports 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 column 2. Columns 4 to 6 again show that the only positive and significant coefficient is the one on *Language less MDI*. Crucially, this result survives the inclusion of colonizer fixed effects, which net out all observable and unobservable characteristics of the colonizing power.

Robustness. Our main results are robust to a range of additional tests, reported in Appendix Table A-16. 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), and to removing from the sample those languages that have gone extinct before 1900 (columns 9-10). Dropping from the sample languages that were colonized by multiple colonial powers leaves our results unchanged, as shown in Table A-17. We obtain similar estimates when employing the Male Dominance Bias from Folklore instead of the MDI, as shown in Table A-18. The MDB has more balanced variation than the MDI in the direction of the male-dominant gap, providing a more direct test of the asymmetric pattern documented above. In Figure A-6, we repeat the bounding exercise of Section 4.5, assigning to every unmatched Indigenous language each possible MDI value, recomputing *Language More MDI* and *Language Less MDI* accordingly, and re-estimating equation 4. The coefficient on *Language Less MDI* remains positive throughout the support and significant at the 5% level except in a narrow interval around 0.77, where it is marginally significant ($p \approx 0.055$).[26] Finally, Table A-19 shows that the coefficient on *Language Less MDI* remains positive and significant when assigning to unmatched languages the MDI of their nearest matched neighbor (columns 2 and 3), or when replacing our tree-based algorithm with a geographic match (columns 4 and 5), as explained in Section 4.5.

5.3 Alternative Competing Explanations

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

[26]In this interval, the assigned MDI approaches the mean MDI of colonizers: as it does so, unmatched languages cross from *Language Less MDI* to *Language More MDI* and mechanically cease to contribute to the coefficient plotted in the figure.

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.

5.3.1 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 5 and 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 significant predictor of extinction even after controlling for group size in addition to our rich set of language-level controls.

5.3.2 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. We therefore 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. We do not find evidence that linguistic distance predicts language extinction. When running a horse-race with the MDI, we find that male dominance remains a powerful explanation for cultural extinction, also in this restricted sample, as shown in Appendix Table A-20.

5.3.3 Distance in Political Complexity

Differences in political complexity between Indigenous groups and their colonizers offer another potential explanation of our results. In a setting close to ours, Cornell and Kalt (1995) find that mismatches between formal post-colonial institutions and pre-colonial informal ones create barriers to economic development among Native American groups. Since differences in gender structure may correlate with distance in pre-colonial institutions, our estimates could in part reflect this institutional channel.

We test this by measuring distance in political complexity, using the Ethnographic Atlas measure, widely used as a proxy for pre-colonial institutions and state capacity (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013; Mayshar *et al.*, 2022). In Table A-21, we find that, on its own, distance in political complexity predicts cultural extinction. However, when distance in political complexity is included alongside 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.

5.3.4 Other Forms of Societal Dominance

A related concern is that distance in male dominance may proxy for distance in other forms of societal dominance—the concentration of political and social power in the hands of a select few. We test this using the Societal Dominance Index (SDI), introduced in Section 3, which captures not only the organizational sophistication of a society but also the degree to which power is concentrated in a dominant group, and is constructed analogously to the MDI.

Table 6 unpacks the role of societal dominance vis-à-vis that of male dominance in shaping cultural extinction. Columns 1 and 2 report horse-races between the MDI and the SDI, both with and without colonizer fixed effects. While both predict extinction, the MDI coefficient remains large and significant, while the SDI coefficient is smaller in magnitude.

We then analyze clashes in societal dominance versus male dominance between Indigenous groups and the colonizers. Column 3 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 4 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.[27] Column 5 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.

5.3.5 The Role of Kinship Tightness

A final alternative explanation for cultural extinction concerns the broader social and moral system within which a society operates, indexed by the tightness of its kinship organization (Enke, 2019). Either kinship regime could plausibly confer advantages in inter-group competition: tight kinship sustains the cohesive in-group coordination useful in direct conflict, while loose kinship sustains the large-scale impersonal cooperation that supports trade, coalition-building, and institutional expansion (Henrich, 2020). The cultural group selection literature establishes that cooperation shapes group success in competition (Henrich, 2004; Richerson *et al.*, 2016), but does not tell us which scale of cooperation—whether in-group or large-scale—matters for cultural survival. This is an open empirical question that we examine in this section using the Kinship Tightness Index (KTI), introduced in Section 3 and constructed analogously to the MDI and SDI.

Table A-22 analyzes clashes in kinship tightness between Indigenous groups and their colonizers, following the same structure as the SDI analysis. Columns 1 and 2 show that the MDI's predictive power survives the inclusion of KTI as a covariate, with or without colonizer fixed effects, while the KTI coefficient remains indistinguishable from zero throughout. Turning to the dyadic specifications, column 3 shows that the absolute distance in KTI is not a significant predictor of cultural extinction. Column 4 decomposes this distance into directional components and finds Indigenous groups with looser kinship

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

than their colonizers face a marginally higher extinction risk. However, when we enter distance in both kinship tightness and male dominance jointly in column 5, this marginal effect fades and our headline asymmetry arises: Indigenous groups that were less male-dominant than their colonizers face a significantly higher extinction risk (at the 5% level).

Taken together, these tests show that male dominance is the dimension of pre-colonial social organization most robustly linked to cultural extinction. Distance in political complexity and distance in societal dominance both predict extinction on their own, but their explanatory power attenuates substantially once distance in male dominance is included, while the male-dominance coefficient remains large and stable. Kinship tightness, despite capturing a moral system theoretically linked to cooperative capacity, yields only a weak and unstable association with extinction risk, and its predictive power does not survive once male dominance distance is accounted for. These findings are consistent with the arguments in Section 2.2 and reinforce the centrality of our proposed mechanism: among the dimensions of pre-colonial social organization we examine, the gender-structure gap between an Indigenous society and its colonizer is the strongest and most robust predictor of cultural survival.

5.4 Mechanisms: Historical Narratives

Why would a larger mismatch in gender structure between colonizers and Indigenous groups predict cultural extinction? Historical and anthropological scholarship points to a common mechanism: European colonization systematically dismantled the gender arrangements of the Indigenous societies they encountered, imposing male-dominant institutions where they had not previously existed. Paula Gunn Allen captures the underlying logic in her account of Indigenous American societies:

> 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.

The specific mechanisms are documented across a range of colonial encounters: French missionaries in New France sought to eliminate polygamy, divorce, and sexual freedom and to install male authority within the family (Devens, 1986); Quaker reformers among the Iroquois replaced women's role as farmers with male-centered agriculture and restructured domestic life around the nuclear family with the husband as its head (Holly, 1990); Jesuits among the Montagnais-Naskapi imposed indissoluble monogamy and refused to engage with anyone but selected male representatives, "thus deforming the Indian structure of governance beyond recognition" (Jaimes and Halsey, 1992); European colonization among the Barí of Colombia introduced "the structure for general male dominance over females" alongside private property and the nuclear family supported by a wage-earning male (Buenaventura-Posso and Brown, 1980); and in colonial Nigeria, the British excluded Yoruba 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 (Oyěwùmí, 1997). Appendix C presents these cases in fuller detail.

These accounts suggest a common pattern: 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 profoundly disruptive the encounter.

5.5 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 language community abandoned its ancestral language, pointing toward assimilation, which may or may not involve population decline.[28] 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 zero otherwise, and *shift to other*, equal to one if the language shifted to a non-colonizer language, and

[28]This field is available for 87% of languages in our colonization sample and contains semi-structured text describing the sociolinguistic behavior of each language, including patterns of bilingualism, domain use, and, crucially for our purposes, language shift. We note that absence of a shift record may reflect incomplete documentation rather than the absence of assimilation.

zero otherwise.

Among Indigenous languages in our sample that exhibit shift based on our indicators, 48% shift toward the colonizer's language, rising to 62% among extinct languages. The remaining 52% shift toward languages other than their assigned colonizers'. Among these, the largest attractor is the Malay-Indonesian cluster (27%), followed by African regional lingua francas including Hausa, Fulfulde, Sudanese and Chadian Arabic, and Swahili (24%). English-lexified creoles such as Tok Pisin and Australian Kriol account for a further 5%.

Table 7 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*.[29] 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 that are 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 to other languages.[30]

6 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 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. In contrast to the colonization exercise, where the analysis is conditioned on already-formed dyads, we leverage a group's exposure to a

[29] 18 observations in our sample shift to both the colonizer's language and a non-colonizer language simultaneously. In these cases, we assign `shift_to_colonizer` = 1 and `shift_to_other` = 0. Results are unchanged when these observations are excluded (Table A-23).

[30] 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.

broader set of potentially competing cultures, i.e., its neighboring groups. 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 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 a highly localized geographical level.

6.1 Measurement and Empirical Setting

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 close to more egalitarian groups. To test for this, we split the African continent into equally-sized grid cells[31], 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 rather than a language, we convert each Murdock group into a language using the "direct links" provided by D-PLACE (see Section 3). This step assigns each of the 843 Murdock groups in the map, and their respective grid cells, to one of 400 Ethnologue languages, and gives us the ancestral distribution of languages across the continent. The modern distribution of languages comes from the Ethnologue map. So, for each grid cell, we know the ancestral language as well as the contemporary one.[32]

We then construct a binary indicator of language survival that equals one if the ancestral language is still spoken today in a given grid cell according to the Ethnologue map, and zero otherwise.[33] Figure 3 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 mostly captures measurement error in the Murdock or Ethnologue map, we validate it in

[31]We leverage the PRIO-GRID dataset, which partitions the world into $0.5^{\circ} \times 0.5^{\circ}$ grids, equivalent to 55 km at the Equator.

[32]Because the Ethnologue lists more than 400 languages for Africa today, we observe more modern languages than ancestral ones. Therefore, a grid cell can be assigned to a modern language that was not present in ancestral times. This is true for 24.3% of cells in our sample. We keep these cells in our baseline analysis, and then show robustness to dropping them.

[33]Table A-5 presents summary statistics for this dataset.

Appendix Table A-24: the grid-cell indicator of language survival is strongly negatively correlated with the language-based extinction indicator from the Ethnologue, and the correlation remains strong and statistically significant across a range of specifications.[34] While measurement error cannot be fully ruled out, these strong correlations suggest that our spatial measure is effective in capturing the contemporary status of languages.

To explore whether neighbor gender structure affects language loss, we leverage within-group variation in (i) language survival and (ii) exposure to male-dominant versus more 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.[35] Crucially, the presence of within-group variation in both language survival and male dominance of the immediate neighbor across grid cells—illustrated in Figure 3 (bottom figure) 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 common to all cells. We estimate:

$$\text{survival}_{igjc} = \theta \text{MDI}_{ij} + \gamma' \mathbf{X}_i + \alpha_g + \delta_c + \epsilon_{igjc} \tag{5}$$

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.[36]

[34] Specifically, the correlation holds with and without country fixed effects (columns 1 and 2), when using the full EGIDS scale as the outcome variable rather than the binary extinction indicator (column 3), and when the data is collapsed to the Murdock-group-by-country level so that language status is regressed on the share of an ethnic homeland's grid cells where the language remains spoken (columns 4–6). In Appendix D, we further show that the grid-cell measure correlates strongly with male dominance, mirroring our findings in Section 4.

[35] When a cell borders multiple groups, we assign to it the MDI of the most male-dominant neighbor. Results are similar using the average MDI across all neighbors. For bordering cells, distance to the closest neighbor is measured as the shortest distance from the cell's centroid to the neighboring group's boundary, consistent with the definition used for non-bordering cells.

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

α_g and δ_c indicate ancestral ethnic homeland and country fixed effects, respectively.[37] The coefficient θ 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, holding all 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, while male-dominant groups are largely insensitive to the gender structure of neighboring groups. To test this, we estimate equation 5 separately for cells belonging to groups with MDI values above and below the sample median.

6.2 Results

Table 8 presents the results. Columns 1–3 report average effects, with the MDI of the closest neighbor entering negatively across all specifications, and significantly when restricting the sample to cells within 20km of the border with the neighboring group (column 2). Column 3 additionally controls for the SDI of the closest neighbor, whose coefficient is close to zero and not statistically significant, and whose inclusion leaves the MDI coefficient unchanged.

The heterogeneous effects by above- versus below-median MDI, reported in columns 4–6 and 7–9, reveal a marked asymmetry consistent with our hypothesis. For more male-dominant groups, the coefficient on neighbor MDI is small and statistically indistinguishable from zero across all specifications, suggesting that these groups are largely insensitive to the gender structure of their neighbors. By contrast, for 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 lower probability of language survival when close to male-dominant neighbors, as compared to cells located close to a more gender-egalitarian neighbor.[38] This confirms that such groups are disproportionately susceptible to the pressures exerted by male-dominant neighbors.

[37]Group size, political complexity and nomadism vary at the ethnic group level. Therefore, they are not included as controls because they are absorbed by group fixed effects. Country fixed effects can be included 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 additional specifications omitting these controls.

[38]Results remain similar when using the average MDI of all closest neighbors rather than the highest (Table A-25), and when dropping the 24.3% of cells containing modern Ethnologue languages that were not present in ancestral times (Table A-26).

6.3 Mechanisms: Military Capacity, Population Pressure, and Expansionism

Our conceptual framework developed in Section 2 identifies two pathways through which male dominance generates expansionist tendencies: (i) the development of military capacity, enabled by the extraction of surplus; and (ii) population pressure, sustained by women's specialization in domestic and reproductive labor. We now provide empirical evidence supporting these mechanisms, leveraging the spatial dataset described in the previous section.

Male dominance and language expansion. We first ask whether languages of male-dominant groups have expanded more than those of gender-equal groups. Our unit of observation for this exercise is a Murdock ethnic group. For each group, we measure its spatial expansion as the difference between the number of grid cells in which its language is spoken today (as per the Ethnologue) and the number of cells its language covered in ancestral times (as per the Murdock map).[39] Whenever there is a one-to-one relationship between Murdock groups and languages, computing the number of grid cells is straightforward. However, as explained above, multiple Murdock groups might share the same language. In these cases, we compute ancestral group size as the total number of cells across all the Murdock groups that share that language. Therefore, Murdock groups with shared languages will also share their expansion measure. We censor negative differences at zero, since our aim is to understand the mechanisms underlying expansion, and take the logarithm of one plus this variable.

Table 9 reports estimates from regressing this expansion measure on the group's MDI and SDI, controlling for our baseline covariates. The MDI is strongly and significantly correlated with expansion across all specifications, with beta coefficients of 0.26 without controls (column 1) and 0.35 with the full set of baseline controls (column 2). The SDI also enters positively in the specification when included alone (column 3), but its coefficient loses statistical significance once the MDI is included (column 4), while the MDI coefficient remains large and significant. This pattern is consistent with our argument that male dominance, rather than hierarchical organization per se, is the more fundamental driver of expansion.

The two underlying channels. Our framework points to two complementary channels

[39] The Ethnologue map, unlike the Murdock map, can have overlapping polygons in which multiple languages coexist; we count all cells in which a language is spoken, regardless of whether it is shared with other languages.

through which male dominance generates expansion: military capacity and population pressure. We measure each at the Murdock-group level. We follow Michalopoulos and Xue (2021) and use the frequency of Folklore-based military- and fertility-related motifs.[40] These measures reflect the cultural salience of military capacity and fertility in a society.

Columns 1–6 of Table 10 examine whether the MDI correlates with these channels. The MDI is strongly and positively associated with military capacity (columns 1 and 3) and with the frequency of fertility motifs in oral tradition (columns 4 and 6). All coefficients are significant at the 1% level and remain so when the SDI is included as a control. An interesting and conceptually meaningful contrast emerges with the SDI: it correlates significantly with military capacity (columns 2–3) but not consistently with fertility (columns 5–6). This asymmetry reflects the conceptual distinction developed in Section 2: while hierarchical organization in general can support the extraction of surplus and the development of military capacity, population pressure is specifically tied to the gender division of labor and women's domestic and reproductive socialization. It is therefore a feature of male dominance in particular, not of hierarchical societies in general.

From channels to expansion. Having shown that male dominance is correlated with both military capacity and fertility, we now ask whether these channels, in turn, predict spatial expansion. Columns 7-9 of Table 10 show that both do. Military capacity enters positively both when included alone (column 7) and when combined with fertility (column 9). Fertility motifs (column 8) also positively predict expansion, and the coefficient is robust to the inclusion of military capacity. The robustness of these effects to joint inclusion suggests that the two channels operate as complementary forces. This is consistent with the idea that male-dominant groups expand both because they have the ability to do so due to their military capacity, and because they have greater demographic incentives to seek new territory due to population pressure. This complementarity likely also accounts for why male dominance consistently outperforms social dominance as a predictor of cultural extinction throughout our analysis: the SDI captures only one of these two pathways (military capacity), while the MDI activates both.

[40]Specifically, for military capacity, we use the Folklore database's *military_related*, *weapon_related*, and *soldier_related* concepts. Because tags capturing similar concepts tend to overlap (a motif tagged *military_related* is likely to also be tagged *weapon_related*), we take the maximum count across these tags for each group to avoid double-counting, and take the logarithm of one plus this variable. For fertility, we use the number of *fertility_related* concepts, and take the logarithm of one plus this variable.

7 Conclusion

This paper examines the determinants of cultural extinction, proxied by language loss, with a focus on the role of societal gender structure. We assemble a novel global dataset combining language status with measures of male dominance in the associated group, alongside geo-climatic and historical characteristics. Our analysis establishes gender structure as a strong and robust predictor of language extinction. Yet, cultural survival is not determined in a vacuum: 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 due to higher military capacity and demographic pressure. 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.

Exploiting local variation in the gender structure of neighboring groups across Africa, we uncover similar patterns: gender-egalitarian societies face heightened risks of cultural erosion when in close proximity with male-dominant neighbors. Using a novel measure of language expansion that compares historical homeland boundaries with contemporary language distributions, we shed light on the mechanisms behind male-dominant expansion. We show that male dominance is strongly correlated with military capacity (proxied by military-related motifs in oral traditions) and higher fertility (proxied by a folklore-based measure), and that both channels in turn predict language expansion.

Our findings contribute to our understanding of cultural evolution by shedding new light on the process of inter-group cultural competition: when groups meet, asymmetries in gender structure are a powerful determinant of which culture prevails. In the modern era of human history we study, societal dominance conferred a comparative advantage in expansionism: the resources extracted from the subordinate group enabled the development of military capacity. Male dominance, when compared to other social dominance structures, enhanced both military capacity—through less costly norm enforcement—and the incentives for territorial expansion—through higher fertility enabled by sexual division of labor—at the expense of more gender-egalitarian groups.

Yet, traits that conferred comparative advantage under one set of conditions need not remain advantageous when those conditions change. A growing literature documents that cultural norms can become maladaptive following such shifts (Gelfand, 2021; Nunn, 2022). Male dominance may represent a particularly relevant case in the present day. The constraints male-dominant systems place on women's economic and political participation may become increasingly costly, as societies confront global challenges such as climate change that require the full mobilization of human capital and inclusive institutions (Bandyopadhyay *et al.*, 2026; Jagnani and Mahadevan, 2025). Whether male dominance has become maladaptive is an empirical question that we leave to future research.

References

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

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

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.

Almeida, D. A. (1997). The hidden half: A history of Native American women's education. *Harvard Educational Review*, **67** (4), 757–772.

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.

Anderson, S. (2025). Gender biased norms. *Available at SSRN 5460014*.

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.

Bandyopadhyay, S., Dutta, P. P., Hari, N. and Maity, B. (2026). Female legislators and forest conservation in india. *Journal of Economic Behavior & Organization*, **241**, 107375.

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. (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 — (2026). *Economic models of identity formation and cultural transmission*. Tech. rep., National Bureau of Economic Research.

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

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.

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 colombia. *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.

Butzer, K. W. (2012). Collapse, environment, and society. *Proceedings of the National Academy of Sciences*, **109** (10), 3632–3639.

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

Chafetz, J. S. (1984). *Sex and Advantage: A Comparative, Macro-Structural Theory of Sex Stratification*. Totowa, NJ: Rowman & Allanheld.

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

— and — (2000). Where's the glue? institutional and cultural foundations of american indian economic development. *The Journal of Socio-Economics*, **29** (5), 443–470.

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

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

Devens, C. (1986). Separate confrontations: Gender as a factor in Indian adaptation to European colonization in New France. *American Quarterly*, **38** (3), 461–480.

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*. New York: Penguin.

D'Amato, M. and Russo, F. F. (2026). *Ancestral Cultural Traits, Colonialism, and its Legacy*. Tech. rep., Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.

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

Enke, B. (2019). Kinship, cooperation, and the evolution of moral systems. *The Quarterly Journal of Economics*, **134** (2), 953–1019.

Fernández, R. (2025). *Understanding cultural change*. Tech. rep., National Bureau of Economic Research.

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.

Gelfand, M. J. (2021). Cultural evolutionary mismatches in response to collective threat. *Current Directions in Psychological Science*, **30** (5), 401–409.

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.

— (2026). Sticky traditions: Origin, persistence, and evolution of cultural norms.

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

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.

Green, R. (1992). *Women in American Indian Society*. Indians of North America, New York: Chelsea House Publishers.

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.

— (2020). *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*. Farrar, Straus and Giroux.

— 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.

Holly, M. (1990). Handsome lake's teachings: The shift from female to male agriculture in Iroquois culture. an essay in ethnophilosophy. *Agriculture and Human Values*, **7** (3–4), 80–94.

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

Jagnani, M. and Mahadevan, M. (2025). Women leaders improve environmental outcomes: Evidence from crop fires in india. *Journal of Public Economics*, **248**, 105443.

Jaimes, M. A. and Halsey, T. (1992). American Indian women: At the center of indigenous resistance in North America. In M. A. Jaimes (ed.), *The State of Native America: Genocide, Colonization, and Resistance*, Boston: South End Press, pp. 311–344.

Jha, S. (2013). Trade, institutions, and ethnic tolerance: Evidence from south asia. *American Political Science Review*, **107** (4), 806–832.

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.

Leacock, E. B. (1981). *Myths of Male Dominance: Collected Articles on Women Cross-Culturally*. New York: Monthly Review Press.

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

Lewis, M. P. (2014). Ethnologue: Languages of the world.

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

Lowes, S. (2024). Culture in historical political economy. In J. A. Jenkins, J. Rubin, J. A. Jenkins and J. Rubin (eds.), *The Oxford Handbook of Historical Political Economy*, Oxford University Press.

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.

Matavelli, I., Grosjean, P., De Haas, R. and Baranov, V. (2026). Masculinity norms and their economic implications. *Annual Review of Economics*, **18**.

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.

McGuirk, E. F. and Nunn, N. (2025). Transhumant pastoralism, climate change, and conflict in africa. *Review of Economic Studies*, **92** (1), 404–441.

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

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

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

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. (2022). On the causes and consequences of cross-cultural differences. *Handbook of Advances in Culture and Psychology*, **9**, 125.

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

— and WANTCHEKON, L. (2011). The slave trade and the origins of mistrust in africa. *American Economic Review*, **101** (7), 3221–3252.

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.

PORTMAN, T. A. A. and HERRING, R. D. (2001). Debunking the Pocahontas paradox: The need for a humanistic perspective. *Journal of Humanistic Counseling, Education and Development*, **40** (2), 185–199.

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

RICHERSON, P., BALDINI, R., BELL, A. V., DEMPS, K., FROST, K., HILLIS, V., MATHEW, S., NEWTON, E. K., NAAR, N., NEWSON, L. *et al.* (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, **39**, e30.

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.

SANDAY, P. R. (1981). *Female Power and Male Dominance: On the Origins of Sexual Inequality*. Cambridge: Cambridge University Press.

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

SPACK, R. (1997). Re-visioning Sioux women: Zitkala-Ša's revolutionary American Indian stories. *Legacy: A Journal of American Women Writers*, **14** (1), 25–42.

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.

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.

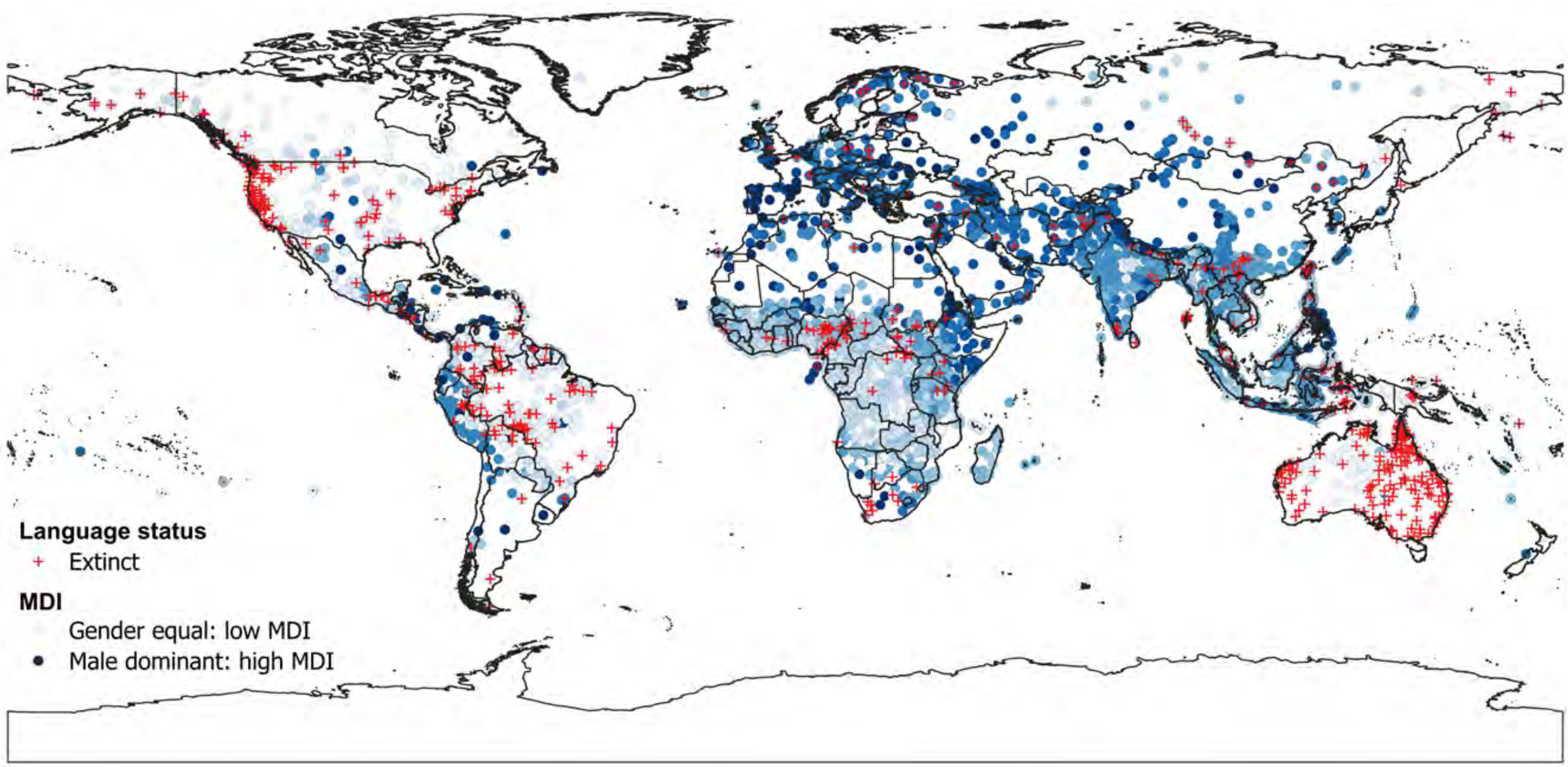
Turchin, P., Currie, T. E., Turner, E. A. L. and Gavrilets, S. (2013). War, space, and the evolution of old world complex societies. *Proceedings of the National Academy of Sciences*, **110** (41), 16384–16389.

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

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

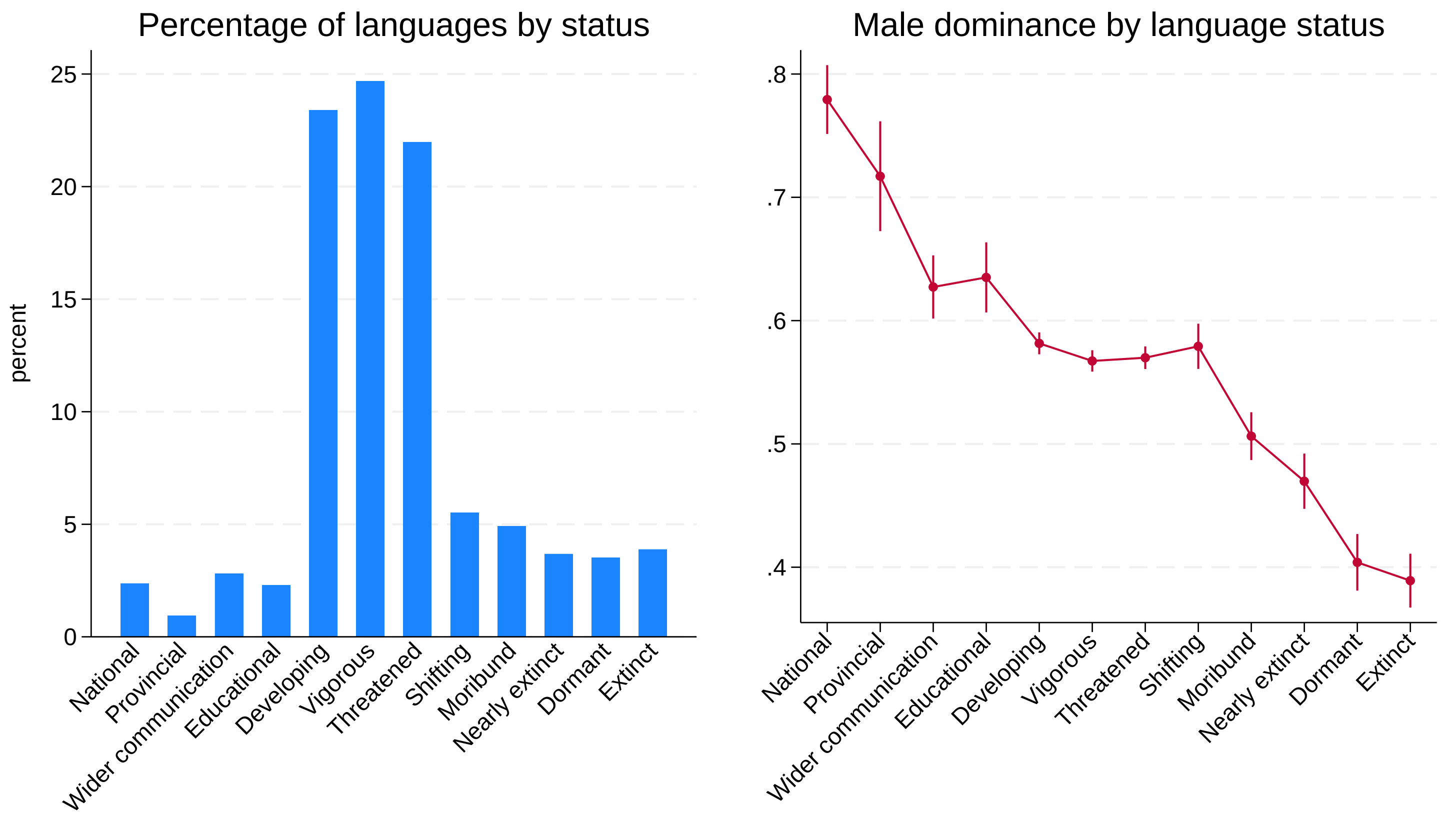
Figures and Tables

FIGURE 1: *Male Dominance and Language Extinction*



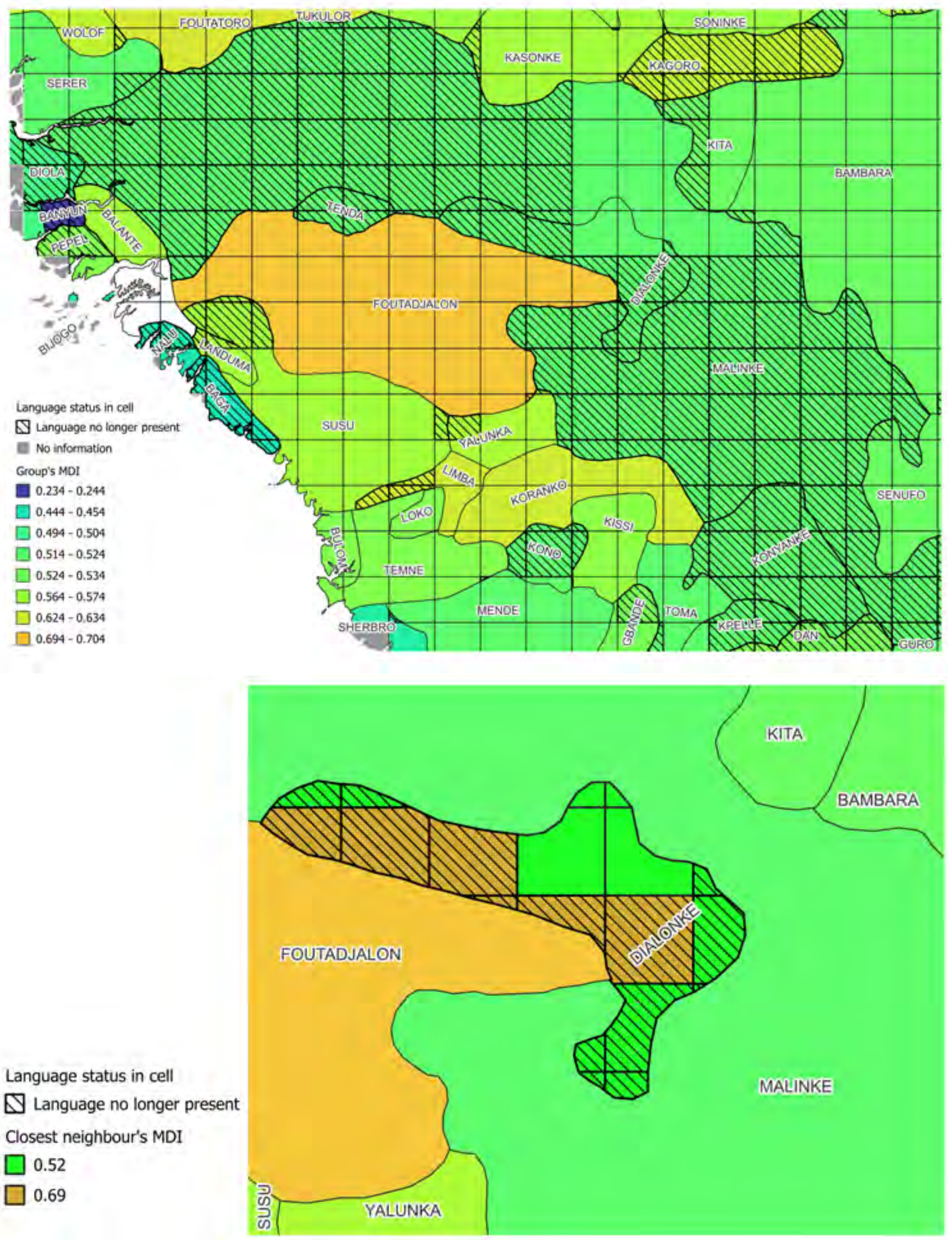
Notes: Each dot on the map represents a language in a country. Darker shades of blue indicate higher degrees of male dominance of the society speaking the language based on Guarnieri & Tur-Prats' (2023) Male Dominance Index (MDI). Red crosses indicate extinct languages based on our definition (i.e., nearly extinct, dormant, or fully extinct in the Ethnologue's EGIDS scale).

Figure 2: *Summary Statistics*



Notes: The histogram on the left displays the distribution of languages across categories in the EGIDS scale. The graph on the right plots the mean of the Male Dominance Index within each language-status category, obtained from a regression of MDI on a full set of status indicators without a constant. Vertical bars are 95 percent confidence intervals.

Figure 3: *Cultural Extinction in Space: Illustration from West Africa*



Notes: Top figure: 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. Bottom figure: 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.

Table 1: *Male Dominance and Language Extinction*

	Dependent variable: extinct (0-1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MDI	-0.47**	-0.53***	-0.53***	-0.53***	-0.51***	-0.51***	-0.37**
	(0.225)	(0.192)	(0.182)	(0.179)	(0.178)	(0.180)	(0.143)
Mean dep. var.	0.106	0.106	0.106	0.106	0.106	0.106	0.106
Observations	5891	5891	5891	5891	5891	5891	5891
Clusters	212	212	212	212	212	212	212
Adjusted R^2	0.076	0.152	0.167	0.168	0.169	0.169	0.202
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 0.22 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***	-0.23***	-0.24***	-0.23***	-0.22***	-0.22***	-0.18**
	(0.091)	(0.083)	(0.082)	(0.080)	(0.079)	(0.079)	(0.079)
Mean dep. var.	0.107	0.107	0.107	0.107	0.107	0.107	0.107
Observations	5854	5854	5854	5854	5854	5854	5854
Clusters	175	175	175	175	175	175	175
Adjusted R^2	0.292	0.300	0.301	0.302	0.302	0.302	0.304
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 0.10 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	-0.33**	-0.03	-0.33**	0.03	0.05
	(0.141)	(0.212)	(0.148)	(0.237)	(0.258)
Absolute distance in the MDI		0.30**			
		(0.149)			
Language more MDI			-0.02		-0.30
			(0.775)		(0.785)
Language less MDI				0.36**	0.37**
				(0.167)	(0.177)
Mean dep. var.	0.115	0.115	0.115	0.115	0.115
Observations	4339	4339	4339	4339	4339
Clusters	106	106	106	106	106
Adjusted R^2	0.257	0.259	0.256	0.259	0.259
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	0.41**	0.05	0.39*	0.05	0.05	
	(0.199)	(0.260)	(0.199)	(0.260)	(0.258)	
Absolute distance in the MDI		0.33**				
		(0.144)				
Language more MDI			-0.94		-0.25	-0.15
			(0.700)		(0.700)	(0.697)
Language less MDI				0.33**	0.32**	0.30**
				(0.142)	(0.143)	(0.142)
Mean dep. var.	0.115	0.115	0.115	0.115	0.115	0.115
Observations	4339	4339	4339	4339	4339	4339
Clusters	106	106	106	106	106	106
Adjusted R^2	0.240	0.259	0.241	0.259	0.259	0.264
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***		-0.01***		-0.01**		-0.01**
	(0.007)	(0.017)	(0.007)		(0.005)		(0.006)		(0.006)
Language's MDI				-0.33**	-0.26**	-0.30**	-0.23*		
				(0.143)	(0.127)	(0.137)	(0.120)		
Language more MDI								-0.10	-0.13
								(0.690)	(0.700)
Language less MDI								0.30**	0.23*
								(0.145)	(0.127)
Mean dep. var.	0.528	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.113
Observations	4279	4279	4279	4279	4279	4279	4279	4279	4279
Clusters	106	106	106	106	106	106	106	106	106
Adjusted R^2	0.235	0.085	0.252	0.257	0.264	0.267	0.274	0.266	0.273
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: *The Role of Societal Dominance*

	Dependent variable: extinct (0-1)				
	(1)	(2)	(3)	(4)	(5)
Language's SDI	-0.12**	-0.11***			
	(0.048)	(0.041)			
Language's MDI	-0.38**	-0.34**			
	(0.174)	(0.162)			
Absolute distance in SDI			0.28***		
			(0.096)		
Language more SDI				0.42	0.44
				(0.274)	(0.269)
Language less SDI				0.29***	0.17***
				(0.105)	(0.064)
Language more MDI					0.72
					(0.701)
Language less MDI					0.36**
					(0.172)
Mean dep. var.	0.121	0.121	0.121	0.121	0.121
Observations	3950	3950	3950	3950	3950
Clusters	102	102	102	102	102
Adjusted R^2	0.228	0.240	0.224	0.224	0.244
Language-level controls (subset)	✓	✓	✓	✓	✓
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 include all the baseline controls, with the exception of pre-colonial institutions (nomadism and political complexity). Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: *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	0.30	0.32					
	(0.303)	(0.356)					
Language more MDI	-1.29**	-1.51**	-1.02***	-1.19**	-0.79**	0.10	0.17
	(0.559)	(0.689)	(0.381)	(0.463)	(0.377)	(0.393)	(0.441)
Language less MDI	0.59**	0.69**	0.31***	0.37***	0.28**	-0.06	-0.07
	(0.271)	(0.302)	(0.109)	(0.140)	(0.112)	(0.046)	(0.046)
Mean dep. var.	0.085	0.097	0.085	0.097	0.068	0.092	0.106
Observations	4339	3788	4339	3788	4339	4339	3788
Clusters	106	99	106	99	106	106	99
Adjusted R^2	0.330	0.424	0.333	0.426	0.307	0.048	0.060
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 (87% 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 8: *Male Dominant Neighbors and Language Survival in Space*

	Dependent variable: language survival (0-1)								
	Average effect			Groups with above-median MDI			Groups with below-median MDI		
		dist<20km			dist<20km			dist<20km	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MDI of closest neighbor	-0.11	-0.17**	-0.11	0.09	0.05	0.08	-0.28***	-0.37***	-0.29***
	(0.086)	(0.085)	(0.090)	(0.149)	(0.124)	(0.156)	(0.096)	(0.110)	(0.081)
SDI of closest neighbor			0.06			0.05			0.09*
			(0.045)			(0.087)			(0.049)
Mean dep. var.	0.482	0.444	0.473	0.544	0.496	0.537	0.436	0.401	0.424
Observations	14936	8540	11860	6668	3834	5222	7270	4055	5796
Clusters	825	816	782	273	322	254	465	410	446
Adjusted R^2	0.552	0.496	0.563	0.577	0.529	0.583	0.518	0.456	0.534
Distance from neighbor	✓	✓	✓	✓	✓	✓	✓	✓	✓
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 neighbor" captures the distance, in degrees, between each grid cell and the closest neighboring 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 absolute latitude, 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. The median MDI for cells in the whole sample (columns 1, 3, 4, 6, 7, 9) is 0.599, and the median MDI for cells in the dist<20km sample (columns 5 and 8) is 0.586. Robust standard errors clustered at the Murdock ethnic homeland level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: *Male Dominance and Expansionism*

	ln(1+expanded cells)			
	(1)	(2)	(3)	(4)
MDI	3.29***	4.50***		3.73***
	(0.595)	(0.584)		(0.635)
SDI			0.65***	0.36
			(0.237)	(0.224)
Beta coefficient	0.256	0.351		0.294
			0.102	0.057
Mean dep. var.	1.161	1.160	1.069	1.063
Observations	744	733	604	592
Adjusted R^2	0.064	0.142	0.069	0.124
Baseline controls		✓	✓	✓

Notes: The unit of observation is a Murdock ethnic group. "ln(1+expanded cells)" measures the spatial expansion of a Murdock ethnic group. "expanded cells" captures the difference between the number of grid cells in which the group's language is currently spoken and the number of grid cells in its ancestral homeland if the difference is positive, and 0 otherwise. "Baseline controls" include the following variables: absolute latitude, the CSI (pre-1500), distance to the coast, elevation, terrain ruggedness, climate variability, malaria prevalence, the share of a group's cells in tropical climate zones, the share of group's cells with any natural resources, the logarithm of the number of natural disasters, and the logarithm of the number of battles. Robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: *Male Dominance, Military Capacity, Fertility, and Expansionism*

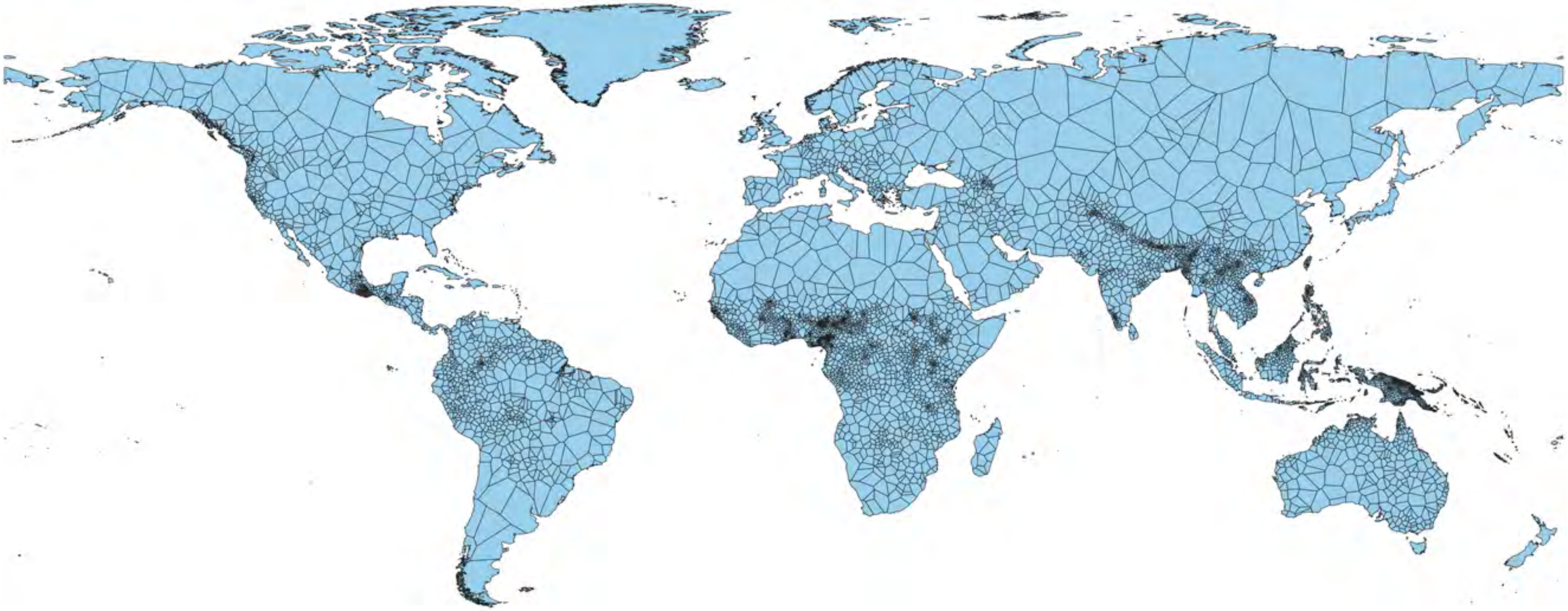
	ln(1+military motifs)			ln(1+fertility motifs)			ln(1+expanded cells)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MDI	0.46***		0.58***	0.29***		0.26***			
	(0.115)		(0.125)	(0.060)		(0.066)			
SDI		0.43***	0.37***		0.08*	0.05			
		(0.061)	(0.064)		(0.040)	(0.041)			
ln(1+military motifs)							0.37**		0.26*
							(0.146)		(0.140)
ln(1+fertility motifs)								1.55***	1.48***
								(0.347)	(0.339)
Beta coefficient	0.157		0.189	0.176		0.163	0.086		0.059
		0.270	0.233		0.094	0.065		0.197	0.187
Mean dep. var.	0.262	0.258	0.256	0.074	0.063	0.064	1.182	1.182	1.182
Observations	713	586	574	713	586	574	790	790	790
Adjusted R^2	0.107	0.180	0.195	0.233	0.104	0.115	0.079	0.102	0.104
Baseline and Folklore controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The unit of observation is a Murdock ethnic group. "ln(1+military motifs)" measures the frequency of military capacity-related motifs in a group's oral tradition, defined as the logarithm of one plus the maximum number of the following Folklore (Michalopoulos and Xue, 2021) concepts: military_related, weapon_related, and soldier_related. "ln(1+expanded cells)" measures the spatial expansion of a Murdock ethnic group. "expanded cells" captures the difference between the number of grid cells in which the group's language is currently spoken and the number of grid cells in its ancestral homeland if the difference is positive, and 0 otherwise. "ln(1+fertility motifs)" measures the frequency of fertility-related motifs in a group's oral tradition, defined as the logarithm of one plus the number of the following Folklore (Michalopoulos and Xue, 2021) concept: fertility_related. "Baseline Controls" includes the following variables: absolute latitude, the CSI (pre-1500), distance to the coast, elevation, terrain ruggedness, climate variability, malaria prevalence, the share of cells in tropical climate zones, the share of cells with any natural resources, the logarithm of the number of natural disasters, and the logarithm of the number of battles. "Folklore Controls" includes the log of the year of the earliest publication Berezkin accessed for each group, and the log number of sources in a group's oral tradition. Robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix (Intended for Online Publication)

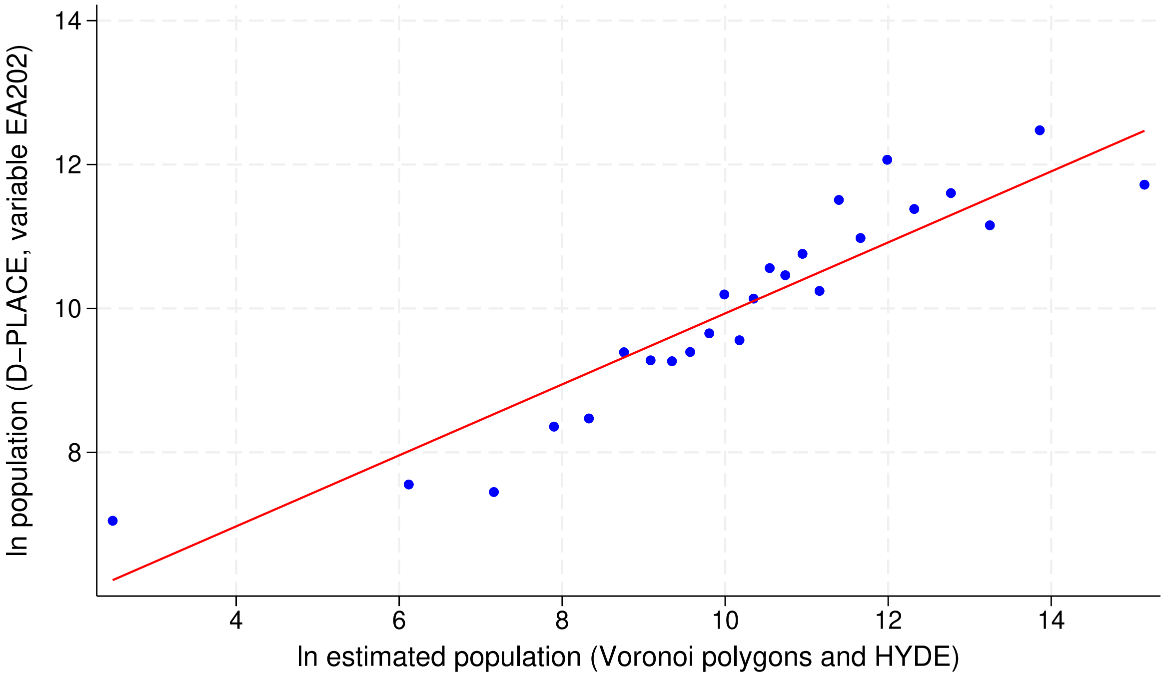
A Additional Figures and Tables

Figure A-1: *Language Voronoi Polygons*



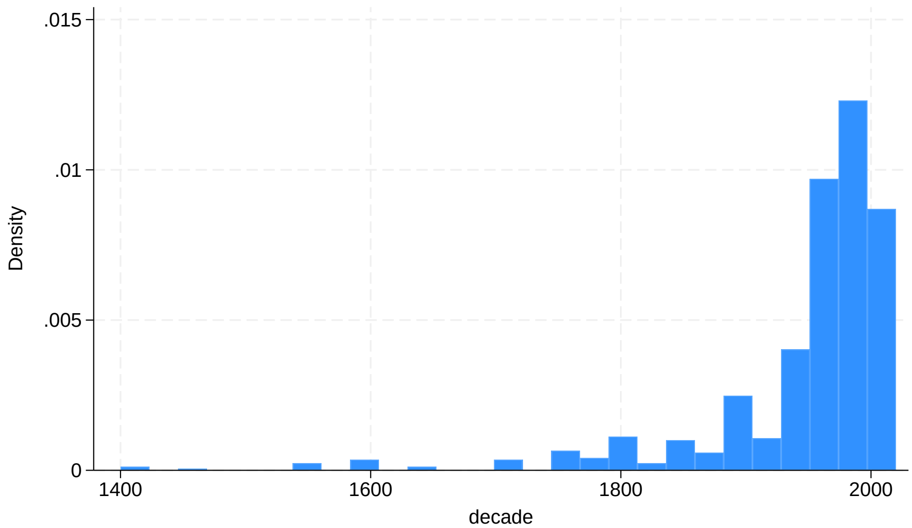
Notes: This map reports Voronoi polygons built around the geographic location (latitude and longitude) of 7,048 living and extinct Indigenous languages and clipped to the coastline; the tessellation is computed separately for each contiguous landmass, so that a landmass is divided only among the languages located on it, and landmasses containing no sample language are assigned to the geographically nearest language. 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-2: *Estimated Population at First Colonial Contact: Measure Validation*



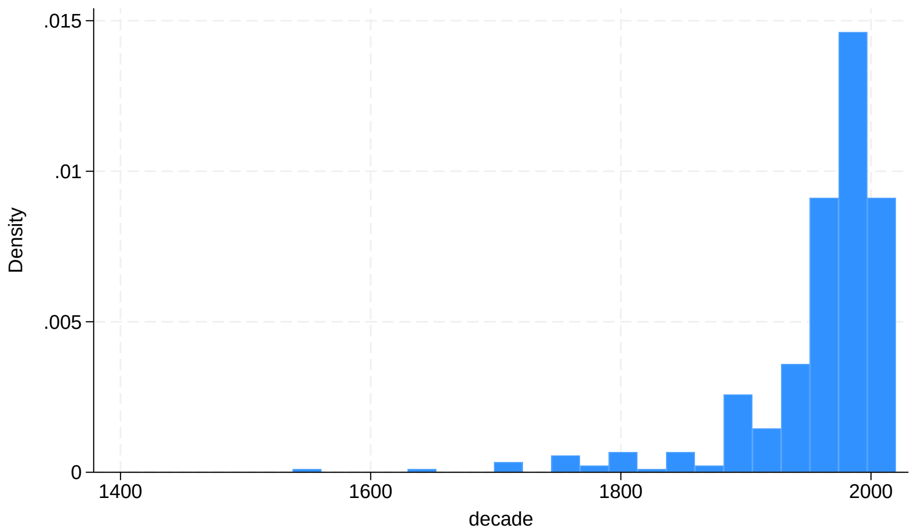
Notes: The figure shows the binned scatterplot of the natural log of estimated population derived from Voronoi polygons combined with HYDE data (x-axis) against the natural log of population reported in the D-PLACE database (EA202 variable, y-axis) for a sample of 683 Indigenous groups-countries. The correlation between the two variables is 0.48 (p-value < 0.001).

Figure A-3: *Language Extinction by Decade: Full Ethnologue Sample*



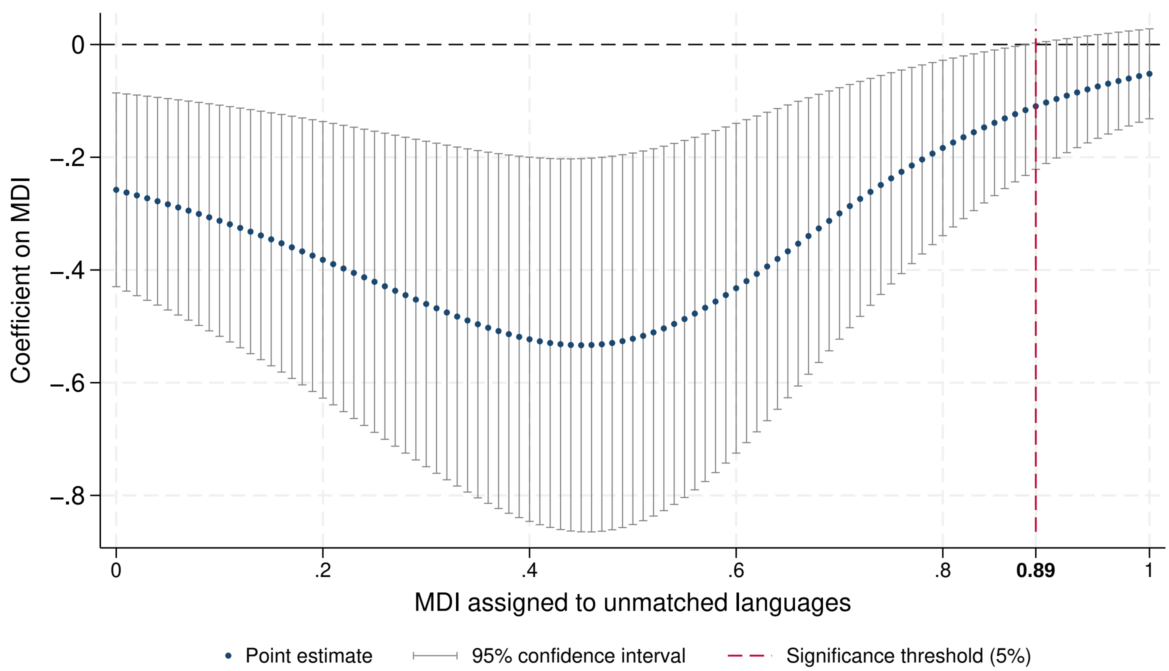
Notes: Based on information on 736 dormant or extinct language-countries (or 721 distinct languages) for which the Ethnologue provides information on the timing of extinction. Dates are approximate and based on when the last known speaker died.

Figure A-4: *Language Extinction by Decade: Estimation Sample*



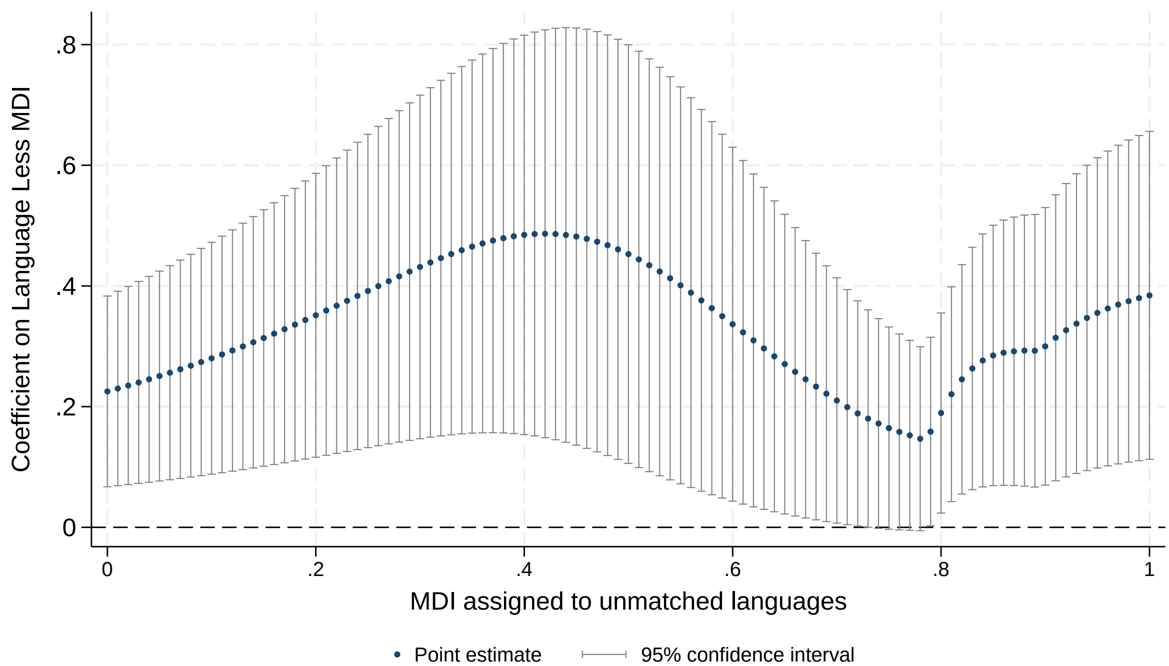
Notes: Based on information on 387 dormant or extinct language-countries (or 384 distinct languages) in our estimation sample for which the Ethnologue provides information on the timing of extinction. Dates are approximate and based on when the last known speaker died.

Figure A-5: *Bounding Exercise: Robustness to Assigning an MDI to Unmatched Languages*



Notes: The figure reports the coefficient on the MDI from estimating equation 1, in which every language that our tree-based procedure fails to match to the Ethnographic Atlas is assigned the value of the MDI shown on the horizontal axis. Each point is a separate regression, estimated on a sample of 8,381 language-countries (or 7,146 distinct languages); vertical bars are 95% confidence intervals from standard errors clustered at the country level as in our baseline specification. The vertical red dashed line marks 0.89, the lowest assigned MDI at which the estimate is no longer significant at the 5% level. The specification excludes nomadism and political complexity from the baseline controls, as both are constructed through the same procedure as the MDI and are therefore unavailable for precisely the languages this exercise is designed to reintroduce.

Figure A-6: *Bounding Exercise: Robustness to Assigning an MDI to Unmatched Languages (European Colonization)*



Notes: The figure reports the coefficient on Language Less MDI$_{ic}$ from estimating equation 4, i.e., regressing language extinction on Language More MDI$_{ic}$, Language Less MDI$_{ic}$ and baseline controls, absorbing colonizer fixed effects and restricting the sample to Indigenous languages, in which every language that our tree-based procedure fails to match to the Ethnographic Atlas is assigned the value of the MDI shown on the horizontal axis; Language More MDI$_{ic}$ and Language Less MDI$_{ic}$ are recomputed at each value from the assigned MDI and the colonizer's MDI. Each point is a separate regression, estimated on a sample of 6,493 languages; vertical bars are 95% confidence intervals from standard errors clustered at the country level. The specification excludes nomadism and political complexity from the baseline controls, as both are constructed through the same procedure as the MDI and are therefore unavailable for precisely the languages this exercise is designed to reintroduce.

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: *Estimated Population at Colonial Contact: Measure Validation*

	Dep. var.: ln population (D-PLACE)		
	(1)	(2)	(3)
ln estimated population	0.49***	0.34***	0.34***
	(0.030)	(0.032)	(0.037)
Observations	895	895	841
Adjusted R^2	0.233	0.427	0.511
Continent FE		✓	
Country FE			✓

Notes: The unit of observation is an Indigenous group in a country. The dependent variable is the natural log of EA202 from the D-PLACE database, measuring the total population of the group. The independent variable is our estimated group population, constructed by overlaying the group's language Voronoi polygon (see Figure A-1) 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" in D-PLACE). Robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-3: *Main Sample: Summary Statistics*

	Mean	St. dev.	Min	Max	N
Extinct	.1062638	.3082011	0	1	5891
MDI	.5622835	.1803991	0	1	5891
Absolute latitude	17.61575	14.54959	.0013047	72.44424	5891
Caloric suitability index	1283.693	739.9183	0	5046.212	5891
Distance to the coast	402.6635	416.8427	.0017361	2551.59	5891
Elevation	672.6201	767.2499	-5.864119	5670.357	5891
Ruggedness	194.1498	215.6894	0	1612.221	5891
Climate variability	-.0362164	.2695868	-1.18	1.09	5891
Malaria index	7.013271	9.149767	0	36.61673	5891
Tropical climate	.5543003	.4718287	0	1	5891
Any natural resource	.442879	.4967686	0	1	5891
Natural disasters (ln)	.7668918	1.290757	0	8.446556	5891
Historical battles (ln)	.1201664	.5165379	0	5.971262	5891
Nomadic	.1226615	.3130394	0	1	5891
Political complexity	1.24897	1.199637	0	4	5891

Table A-4: *Colonization Sample: Summary Statistics*

	Mean	St. dev.	Min	Max	N
Extinct	.1145425	.3185057	0	1	4339
MDI	.5261063	.1737354	0	.955752	4339
Absolute distance MDI	.269255	.1766642	.0009971	.8259152	4339
More male dominant	.0023534	.0129129	0	.2112644	4339
Less male dominant	.2669016	.1797221	0	.8259152	4339
Absolute latitude	14.18372	11.91734	.0013047	69.87904	4339
Caloric suitability index	1201.761	705.9773	0	5046.212	4339
Distance to the coast	376.4746	397.4931	.0103727	1808.291	4339
Elevation	608.0278	696.5622	-5.864119	5670.357	4339
Ruggedness	175.1001	201.5787	0	1612.221	4339
Climate variability	-.037684	.2641643	-1.18	1.09	4339
Malaria index	8.249067	9.499574	0	36.61673	4339
Tropical climate	.6430124	.4492251	0	1	4339
Any natural resource	.4385803	.4962705	0	1	4339
Natural disasters (ln)	.6867222	1.19502	0	6.942157	4339
Historical battles (ln)	.0554746	.319325	0	4.553877	4339
Nomadic	.1237849	.3182704	0	1	4339
Political complexity	1.060888	1.100173	0	4	4339
SDI	.3223869	.2523384	0	1	3950
KTI	.7407705	.2794048	0	1	4116
Population size at contact (ln)	8.833591	2.441613	0	15.92042	4279

TABLE A-5: *Spatial Analysis Sample: Summary Statistics*

	Mean	St. dev.	Min	Max	N
Language survival	.4822576	.4997018	0	1	14936
Ancestral MDI	.6248653	.1544999	.1835572	.9662781	13940
MDI of closest neighbor (max)	.6213129	.1532062	.1835572	.9662781	14936
MDI of closest neighbors (avg)	.6270146	.1538075	.1835572	.9662781	14936
SDI of closest neighbor (max)	.4537593	.2498811	0	1	12979
SDI of closest neighbor (avg)	.4409143	.2461764	0	1	12979
Distance from neighbor (in km)	31.01218	41.63524	.0366444	435.4493	14936
Group size (log # of grids)	3.523304	.9503053	1.098612	5.537334	14936
Cell crossed by country border	.123594	.3291288	0	1	14936
Absolute latitude	14.15808	9.297755	.0029033	37.1494	14936
Caloric suitability index	875.9737	616.8133	0	2376.317	14931
Distance to the coast	635.1794	465.7325	.0021035	1813.172	14936
Elevation	637.2201	462.5622	-105.6135	3358.53	14936
Ruggedness	70.9378	90.89696	0	1274.637	14936
Climate variability	-.0767064	.2926724	-.87	.61	13632
Malaria index	10.51509	10.14526	0	38.08099	14925
Tropical climate (binary)	.4716792	.499214	0	1	14936
Any natural resource	.1261382	.3320162	0	1	14936
Natural disasters (ln)	.1936752	.5404557	0	4.127134	14936
Historical battles (ln)	.0044046	.0693981	0	2.944439	14936
Nomadic	.3125881	.4635644	0	1	14188
Political complexity	1.532753	1.034427	0	4	14014

Table A-6: *Horserace: MDI vs Individual Ethnic Traits*

	Dependent variable: extinct (0-1)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MDI	-0.49**	-0.50***	-0.56***	-0.55***	-0.53***	-0.48**	-0.19*	-0.47**	-0.79***	-0.57***
	(0.215)	(0.164)	(0.193)	(0.185)	(0.183)	(0.200)	(0.102)	(0.206)	(0.206)	(0.192)
Mean dep. var.	0.106	0.106	0.106	0.106	0.106	0.106	0.106	0.106	0.106	0.106
Control:	*plough*	*brideprice*	*polygyny*	*nuclear*	*matrilineal*	*fishing*	*gathering*	*pastoralism*	*non-herding husbandry*	*shifting agriculture*
Observations	5891	5891	5891	5891	5891	5891	5891	5891	5891	5891
Clusters	212	212	212	212	212	212	212	212	212	212
Adjusted R^2	0.169	0.182	0.171	0.170	0.169	0.170	0.231	0.169	0.192	0.182

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-7: *Male Dominance Bias (Michalopoulos and Xue, 2021) and MDI*

	MDI		
	(1)	(2)	(3)
Male Dominance Bias	0.77***	0.56***	0.27***
	(0.191)	(0.158)	(0.091)
Mean dep. var.	0.563	0.563	0.563
Observations	5791	5791	5791
Adjusted R^2	0.081	0.180	0.419
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-8: *Male Dominance Bias and Cultural Extinction*

	Dependent variable: extinct (0-1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male Dominance Bias	-0.21	-0.43*	-0.46*	-0.46*	-0.44*	-0.43*	-0.28*
	(0.240)	(0.258)	(0.257)	(0.254)	(0.246)	(0.243)	(0.172)
Mean dep. var.	0.100	0.100	0.100	0.100	0.100	0.100	0.100
Observations	6733	6733	6733	6733	6733	6733	6733
Clusters	215	215	215	215	215	215	215
Adjusted R^2	0.002	0.075	0.088	0.091	0.098	0.101	0.161
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.

Table A-9: *Additional Ethnic Traits: Patrilocality and Exogamy*

	Dependent variable: extinct (0-1)			
	(1)	(2)	(3)	(4)
MDI	-0.51**	-0.51**	-0.49**	-0.49**
	(0.183)	(0.181)	(0.171)	(0.172)
Patrilocality		0.01		-0.01
		(0.031)		(0.024)
Exogamy			0.03	0.04
			(0.037)	(0.033)
Mean dep. var.	0.103	0.103	0.103	0.103
Observations	5700	5700	5700	5700
Clusters	212	212	212	212
Adjusted R^2	0.163	0.163	0.165	0.165

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-10: *MDI, SDI and KTI: Cross-correlation Table*

Variables	MDI	SDI	KTI
MDI	1.000		
SDI	0.672	1.000	
KTI	-0.375	-0.235	1.000

Table A-11: *Comparison of the Relationship of MDI, SDI, and KTI with Extinction*

	Dependent variable: extinct (0-1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MDI	-0.454**			-0.427**	-0.490***			-0.419**
	(0.209)			(0.200)	(0.182)			(0.164)
SDI		-0.259**		-0.084*		-0.246***		-0.090**
		(0.113)		(0.050)		(0.088)		(0.042)
KTI			0.014	-0.084**			0.057	-0.007
			(0.057)	(0.033)			(0.077)	(0.049)
Mean dep. var	0.103	0.103	0.103	0.103	0.109	0.109	0.109	0.109
Observations	6236	6236	6236	6236	5140	5140	5140	5140
Adjusted R^2	0.080	0.052	0.000	0.091	0.169	0.139	0.104	0.173
Controls					✓	✓	✓	✓

Notes: Controls include all language-level controls except pre-colonial institutions (nomadism and political complexity). Robust standard errors clustered at the country level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-12: *Comparison of the Relationship of MDI, SDI, and KTI with Extinction. Quadratic KTI*

	Dependent variable: extinct (0-1)			
	(1)	(2)	(3)	(4)
KTI	0.234**	0.054	-0.030	-0.178*
	(0.111)	(0.080)	(0.120)	(0.105)
KTI^2	-0.205*	-0.128	0.080	0.156
	(0.107)	(0.085)	(0.125)	(0.110)
MDI		-0.419**		-0.425**
		(0.201)		(0.166)
SDI		-0.086*		-0.090**
		(0.050)		(0.041)
Mean dep. var	0.103	0.103	0.109	0.109
Observations	6236	6236	5140	5140
Adjusted R^2	0.004	0.092	0.104	0.174
Controls			✓	✓

Notes: Controls include all language-level controls except pre-colonial institutions (nomadism and political complexity). Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-13: *Male Dominance and Extinction: Addressing Selection into the Sample*

	Dependent variable: extinct (0-1)				
			dist<100km		dist<100km
	(1)	(2)	(3)	(4)	(5)
MDI (tree-based matching)	-0.37**				
	(0.143)				
MDI (tree-based matching)		-0.25**	-0.31**		
+ nearest-neighbor MDI		(0.116)	(0.140)		
MDI of nearest direct-match				-0.30***	-0.30**
				(0.114)	(0.117)
Mean dep. var.	0.106	0.121	0.112	0.139	0.117
Observations	5891	8381	7219	5859	5357
Clusters	212	229	215	210	203
Adjusted R^2	0.202	0.200	0.187	0.207	0.175
Baseline controls	✓	✓	✓	✓	✓

Notes: The unit of observation is a language in a country. The table compares three ways of assigning the Male Dominance Index to languages. *MDI (tree-based matching)* is our baseline measure, assigned to each language from the societies in the Ethnographic Atlas closest to it in the linguistic tree. *MDI (tree-based matching) + nearest-neighbor MDI* retains the baseline measure wherever it exists and assigns to every remaining language the MDI of its geographically closest tree-matched language; *MDI of nearest direct-match* discards the tree-based assignment and assigns to every language the MDI of the geographically closest language that matches directly to a society in the Atlas. The controls *nomadism* and *political complexity* are constructed in the same way. Columns 3 and 5 restrict geographically assigned observations to those whose donor lies within 100 km. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-14: *Male Dominance and Extinction: Robustness (I)*

	only Indigenous languages				only direct matches		cluster by family		year of observation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
MDI	-0.49**	-0.37**			-0.59**	-0.54***	-0.47**	-0.37***	-0.47**	-0.36**
	(0.239)	(0.148)			(0.240)	(0.166)	(0.215)	(0.134)	(0.222)	(0.138)
Alt. MDI			-0.44**	-0.34**						
			(0.214)	(0.133)						
Mean dep. var.	0.109	0.109	0.106	0.106	0.088	0.088	0.106	0.106	0.106	0.106
Observations	5487	5487	5891	5891	1122	1122	5891	5891	5868	5868
Clusters	199	199	212	212	160	160	95	95	212	212
Adjusted R^2	0.079	0.213	0.072	0.201	0.140	0.251	0.076	0.202	0.084	0.207
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 9 and 10 control for the logarithm of the year in which the Murdock ethnicity was observed by the ethnographer. Robust standard errors clustered at the country level are reported in parentheses, except in columns 7 and 8, where they are clustered at the linguistic family level. *** p<0.01, ** p<0.05, * p<0.1.

Table A-15: *Male Dominance and Extinction: 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	-0.54**	-0.43***	-0.21*	-0.14*	-0.45**	-0.36**	-0.50***	-0.34***
	(0.252)	(0.157)	(0.118)	(0.069)	(0.220)	(0.141)	(0.031)	(0.032)
Mean dep. var.	0.215	0.215	0.039	0.039	0.101	0.101	0.123	0.123
Observations	5891	5891	5891	5891	5855	5855	4794	4794
Clusters	212	212	212	212	212	212		
Adjusted R^2	0.056	0.224	0.037	0.105	0.073	0.192	0.078	0.225
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 extinct or dying (either shifting, moribund, nearly extinct, dormant or extinct 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-16: *European Colonization: Robustness (I)*

Robustness:	controls		robust se		cluster language		alt. MDI		extinct post 1900	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Language's MDI	0.01		0.05		0.05				0.06	
	(0.338)		(0.094)		(0.100)				(0.250)	
Language more MDI	0.34	0.35	-0.30	-0.25	-0.30	-0.25			-0.26	-0.20
	(0.795)	(0.610)	(0.395)	(0.386)	(0.417)	(0.409)			(0.764)	(0.695)
Language less MDI	0.53**	0.51**	0.37***	0.32***	0.37***	0.32***			0.38**	0.32**
	(0.236)	(0.206)	(0.088)	(0.032)	(0.093)	(0.033)			(0.176)	(0.140)
Colonizer's MDI		0.01		0.05		0.05		0.05		0.06
		(0.338)		(0.094)		(0.100)		(0.263)		(0.250)
Language's MDI (alt)							0.05			
							(0.263)			
Language more MDI (alt)							-0.70	-0.66		
							(0.602)	(0.491)		
Language less MDI (alt)							0.34*	0.29**		
							(0.189)	(0.135)		
Mean dep. var.	0.113	0.113	0.115	0.115	0.115	0.115	0.115	0.115	0.110	0.110
Observations	4411	4411	4339	4339	4339	4339	4339	4339	4314	4314
Clusters	106	106			3862	3862	106	106	106	106
Adjusted R^2	0.218	0.218	0.259	0.259	0.259	0.259	0.255	0.255	0.251	0.251
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 language level. Columns 7 and 8 employ an alternative MDI. In columns 9 and 10, 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-17: *European Colonization: Robustness (II)*

	Dependent variable: extinct (0-1)			
	(1)	(2)	(3)	(4)
Language's MDI	-0.36**	-0.04	0.08	
	(0.136)	(0.219)	(0.269)	
Absolute distance in the MDI		0.32*		
		(0.162)		
Language more MDI			-0.72	-0.60
			(0.814)	(0.714)
Language less MDI			0.42**	0.34**
			(0.201)	(0.132)
Mean dep. var.	0.123	0.123	0.123	0.123
Observations	3595	3595	3595	3595
Clusters	79	79	79	79
Adjusted R^2	0.318	0.320	0.321	0.322
Language-level controls	✓	✓	✓	✓
Colonizer fixed effects				✓

Notes: The unit of observation is a language in a country. Languages that were colonized by multiple countries are removed from the sample. Robust standard errors are reported in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1.

TABLE A-18: *European Colonization: Male Dominance Bias (Michalopoulos and Xue, 2021)*

	Dependent variable: extinct (0-1)		
Measure:	MDI (baseline)	Male Dominance Bias	
	(1)	(2)	(3)
Language more male dominant	-0.14	-0.05	0.17
	(0.697)	(0.128)	(0.132)
Language more gender equal	0.30**	0.46*	0.41*
	(0.144)	(0.234)	(0.242)
Mean dep. var.	0.115	0.115	0.115
Observations	4289	4289	4277
Clusters	100	100	88
Adjusted R^2	0.264	0.256	0.360
Language-level controls	✓	✓	✓
Colonizer fixed effects	✓	✓	
Country fixed effects			✓

Notes: The unit of observation is a language in a country. The sample includes only countries that were colonized and only languages indigenous to the country. In column 1, we retain the sample of languages with a non-missing value of the Male Dominance Bias index for comparison. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Magnitudes: in column 1, a one standard deviation increase in *Language more gender equal* (using the MDI) increases extinction by 0.17 standard deviations; in column 2, a one standard deviation increase in *Language more gender equal* (using the MDB) increases extinction by 0.10 standard deviations.

Table A-19: *Male Dominance and Extinction: Addressing Selection into the Sample (European Colonization)*

	Dependent variable: extinct (0-1)				
Matching procedure:	tree-based	tree based + nearest-neighbor		nearest direct match	
			dist<100km		dist<100km
	(1)	(2)	(3)	(4)	(5)
Language more MDI	-0.15	-0.92	-0.59	-0.03	-0.10
	(0.697)	(0.621)	(0.585)	(0.442)	(0.372)
Language less MDI	0.30**	0.20**	0.25**	0.29**	0.27**
	(0.142)	(0.092)	(0.122)	(0.135)	(0.135)
Mean dep. var.	0.115	0.122	0.115	0.144	0.122
Observations	4339	6777	5746	4555	4180
Clusters	106	118	110	106	105
Adjusted R^2	0.264	0.256	0.239	0.264	0.228
Baseline controls	✓	✓	✓	✓	✓
Colonizer fixed effects	✓	✓	✓	✓	✓

Notes: The unit of observation is a language in a country. The sample includes only countries that were colonized and only languages indigenous to the country. The table compares three ways of assigning the Male Dominance Index to languages. The *tree-based* one (column 1) is our baseline measure, assigned to each language from the societies in the Ethnographic Atlas closest to it in the linguistic tree. The *tree-based + nearest-neighbor* one (columns 2 and 3) retains the baseline measure wherever it exists and assigns to every remaining language the MDI of its geographically closest tree-matched language; the *nearest direct match* one (columns 4 and 5) discards the tree-based method and assigns to every language the MDI of the geographically closest language that matches directly to a society in the Atlas. The controls *nomadism* and *political complexity* are constructed in the same way. Columns 3 and 5 restrict geographically assigned observations to those whose donor lies within 100 km. Robust standard errors clustered at the country level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-20: *Interpretation: Distance in Gender Norms vs Linguistic Distance*

	Dependent variable: extinct (0-1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Language's MDI	-0.31**					
	(0.142)					
Absolute distance in the MDI		0.38**		0.37**		
		(0.153)		(0.157)		
Linguistic distance			0.05	0.04		0.04
			(0.039)	(0.042)		(0.039)
Language more MDI					0.36	0.28
					(0.583)	(0.562)
Language less MDI					0.38**	0.37**
					(0.152)	(0.155)
Mean dep. var.	0.038	0.038	0.038	0.038	0.038	0.038
Observations	448	448	448	448	448	448
Adjusted R^2	0.474	0.476	0.460	0.476	0.475	0.475
Language-level controls	✓	✓	✓	✓	✓	✓

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 A-21: *Interpretation: Distance in Gender Norms vs Distance in Political Institutions*

	Dependent variable: extinct (0-1)			
	(1)	(2)	(3)	(4)
Distance in political complexity	0.29**	0.15***		0.07*
	(0.118)	(0.045)		(0.035)
Language more MDI			-0.24	-0.15
			(0.683)	(0.697)
Language less MDI			0.33**	0.30**
			(0.141)	(0.142)
Mean dep. var.	0.115	0.115	0.115	0.115
Observations	4339	4339	4339	4339
Clusters	106	106	106	106
Adjusted R^2	0.086	0.249	0.263	0.264
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. *Distance in political complexity* is the distance between the colonizer's political complexity and those of the Indigenous group, normalized to the $[0, 1]$ interval by subtracting the sample minimum and dividing by the sample range. 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 A-22: *The Role of Kinship Tightness*

	Dependent variable: extinct (0-1)				
	(1)	(2)	(3)	(4)	(5)
Language's KTI	0.01	0.01			
	(0.057)	(0.063)			
Language's MDI	-0.31**	-0.28**			
	(0.135)	(0.132)			
Absolute distance in KTI			0.06		
			(0.080)		
Language more KTI				0.06	0.03
				(0.080)	(0.078)
Language less KTI				0.19*	0.16
				(0.096)	(0.100)
Language more MDI					0.13
					(0.730)
Language less MDI					0.28**
					(0.140)
Mean dep. var.	0.114	0.114	0.114	0.114	0.114
Observations	4116	4116	4116	4116	4116
Clusters	106	106	106	106	106
Adjusted R^2	0.258	0.265	0.252	0.252	0.266
Language-level controls	✓	✓	✓	✓	✓
Colonizer fixed effects		✓	✓	✓	✓

Notes: The unit of observation is a language in a country. The sample includes only countries that were colonized and only languages indigenous to the country. Language-level controls include 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 A-23: *Male Dominance and Language Shift. Robustness*

	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	0.29	0.30					
	(0.292)	(0.343)					
Language more MDI	-1.34**	-1.58**	-1.09***	-1.29***	-0.89**	0.11	0.19
	(0.531)	(0.658)	(0.372)	(0.445)	(0.381)	(0.396)	(0.444)
Language less MDI	0.58**	0.67**	0.30***	0.37***	0.27**	-0.06	-0.06
	(0.262)	(0.292)	(0.108)	(0.138)	(0.111)	(0.046)	(0.045)
Mean dep. var.	0.081	0.093	0.081	0.093	0.066	0.093	0.106
Observations	4321	3770	4321	3770	4321	4321	3770
Clusters	106	99	106	99	106	106	99
Adjusted R^2	0.339	0.434	0.341	0.436	0.313	0.048	0.060
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. We drop the 18 observations for which we document a language shift both to a colonizer's language and to a non-colonizer's language. *Language Use Information* restricts the sample to those observations for which the Ethnologue included language use information (87% 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 A-24: *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***	-1.07***			
	(0.024)	(0.011)	(0.156)			
Share of survived cells				-0.02**	-0.03***	-1.54***
				(0.009)	(0.009)	(0.164)
Mean dep. var.	0.024	0.024	4.822	0.008	0.007	4.812
Observations	15052	15052	14808	1013	1008	987
Clusters	790	790	787	790	788	784
Adjusted R^2	0.026	0.518	0.290	0.005	0.219	0.229
Basic controls		✓	✓		✓	✓
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 EGIDS 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 the grid cell today and 0 otherwise; in the Murdock-group analysis, "share of surviving in cells" captures the share of ancestral grid cells in which the language is still spoken today. *Basic controls* include distance to closest neighbor, an indicator if the cell is split by a country border, and the size of the ethnic homeland, measured via the logarithm of the total number of grids that a Murdock ethnic homeland occupies. Robust standard errors clustered at the Murdock ethnic homeland level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-25: *Male Dominant Neighbors 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		
		dist<20km			dist<20km			dist<20km	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MDI of closest neighbor	-0.09	-0.13	-0.08	0.14	0.16	0.15	-0.29***	-0.42***	-0.31***
	(0.092)	(0.095)	(0.097)	(0.155)	(0.136)	(0.166)	(0.102)	(0.121)	(0.088)
SDI of closest neighbor			0.07			0.04			0.10*
			(0.050)			(0.093)			(0.053)
Mean dep. var.	0.482	0.444	0.473	0.544	0.496	0.537	0.436	0.401	0.424
Observations	14936	8540	11860	6668	3834	5222	7270	4055	5796
Clusters	825	816	782	273	322	254	465	410	446
Adjusted R^2	0.552	0.495	0.563	0.577	0.530	0.583	0.518	0.456	0.534
Distance to neighbor	✓	✓	✓	✓	✓	✓	✓	✓	✓
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. *Other controls* are grid-cell level controls including latitude, 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. The median MDI for cells in the whole sample (columns 1, 3, 4, 6, 7, 9) is 0.599, and the median MDI for cells in the dist<20km sample (columns 5 and 8) is 0.586. Robust standard errors clustered at the Murdock ethnic homeland level are reported in parentheses. Robust standard errors clustered at the Murdock ethnic homeland level are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A-26: *Male Dominant Neighbors 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		
		dist<20km			dist<20km			dist<20km	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MDI of closest neighbor	-0.13	-0.29***	-0.11	0.14	-0.07	0.14	-0.43***	-0.49**	-0.39***
	(0.094)	(0.112)	(0.103)	(0.118)	(0.133)	(0.153)	(0.132)	(0.194)	(0.126)
SDI of closest neighbor			0.02			-0.03			0.09
			(0.050)			(0.084)			(0.059)
Mean dep. var.	0.636	0.589	0.631	0.671	0.622	0.660	0.615	0.566	0.616
Observations	11312	6424	8865	5188	2982	4044	5386	2946	4199
Clusters	768	752	711	235	297	213	451	377	421
Adjusted R^2	0.580	0.523	0.598	0.644	0.581	0.648	0.531	0.469	0.555
Distance to neighbor	✓	✓	✓	✓	✓	✓	✓	✓	✓
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. *Other controls* are grid-cell level controls including latitude, 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. The median MDI for cells in the whole sample (columns 1, 3, 4, 6, 7, 9) is 0.610, and the median MDI for cells in the dist<20km sample (columns 5 and 8) is 0.587. Robust standard errors clustered at the Murdock ethnic homeland level 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-7 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-7: *Example of Linguistic Tree*

Muskogean
- ▼ **Alabaman-Koasati**
  - *Alabama*
  - *Apalachee*
  - *Koasati*
- ▼ **Western Muskogean**
  - *Chickasaw*
  - *Choctaw*
- *Creek*
- *Mikasuki*

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-8 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-8: *Direct Matches between Languages and Murdock Groups*

```
Muskogean
      ▼ Alabaman-Koasati
            Alabama
            Apalachee
            Koasati
      ▼ Western Muskogean
            Chickasaw -> Chickasaw (direct)
            Choctaw -> Choctaw (direct)
      Creek -> Creek (direct)
      Mikasuki -> Seminole (direct)
```

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 to one another. Based on this logic, to unmatched languages we assign the Murdock ethnic group associated with 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-9 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-9: *Assigning Murdock Groups to Nodes in the Linguistic Tree*

```
Muskogean [Creek + Seminole + Chickasaw + Choctaw]
      ▼ Alabaman-Koasati [Ø]
            Alabama
            Apalachee
            Koasati
      ▼ Western Muskogean [Chickasaw + Choctaw]
            Chickasaw -> Chickasaw (direct)
            Choctaw -> Choctaw (direct)
      Creek -> Creek (direct)
      Mikasuki -> Seminole (direct)
```

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-10 below illustrates the procedure for the *Alabama* language. We first move up to the **Alabaman-Koasati** 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 Choctaw (step 3).

FIGURE A-10: *Moving up the Tree to Identify Matches*

Step 1

Muskogean [Creek + Seminole + Chickasaw + Choctaw]
▼ **Alabaman-Koasati** [Ø]
Alabama
Apalachee
Koasati
▼ **Western Muskogean** [Chickasaw + Choctaw]
Chickasaw -> Chickasaw (direct)
Choctaw -> Choctaw (direct)
Creek -> Creek (direct)
Mikasuki -> Seminole (direct)

Step 2

Muskogean [Creek + Seminole + Chickasaw + Choctaw]
▼ ~~**Alabaman-Koasati** [Ø]~~
Alabama
Apalachee
Koasati
▼ **Western Muskogean** [Chickasaw + Choctaw]
Chickasaw -> Chickasaw (direct)
Choctaw -> Choctaw (direct)
Creek -> Creek (direct)
Mikasuki -> Seminole (direct)

Step 3

Muskogean [Creek + Seminole + Chickasaw + Choctaw]
▼ **Alabaman-Koasati** [Ø]
Alabama -> Creek + Seminole + Chickasaw + Choctaw
Apalachee
Koasati
▼ **Western Muskogean** [Chickasaw + Choctaw]
Chickasaw -> Chickasaw (direct)
Choctaw -> Choctaw (direct)
Creek -> Creek (direct)
Mikasuki -> Seminole (direct)

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 language 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*, *Choctaw*, *Creek*, and *Mikasuki*, in this case).

The respective scores are displayed in Figure A-11 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-11: *Assigning a Distance Score to each Match*

```
Muskogean [Creek + Seminole + Chickasaw + Choctaw]
     ▼ Alabaman-Koasati [Ø]
          Alabama -> Creek[2]+Seminole[2]+Chickasaw[3]+Choctaw[3]
          Apalachee
          Koasati
     ▼ Western Muskogean [Chickasaw + Choctaw]
          Chickasaw -> Chickasaw (direct)
          Choctaw -> Choctaw (direct)
     Creek -> Creek (direct)
     Mikasuki -> Seminole (direct)
```

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 *Creek's*, *Seminole's*, *Chickasaw's*, and *Choctaw'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.

To compute weights and obtain the weighted average, we employ the following procedure. Let ℓ be a language and let $M(\ell)$ denote the set of Murdock groups to which it can be matched, indexed by i, with $d_{i\ell}$ denoting the distance score of match i. We invert these distances so that proximity is rewarded, rather than remoteness. Letting $D_\ell = \max_{i \in M(\ell)} d_{i\ell}$ be the largest distance recorded for that language, we invert the distances by setting

$$\tilde{w}_{i\ell} = (D_\ell - d_{i\ell}) + 1,$$

and normalize each of these by the row-sum of the inverted weights, so that the weights within a language sum to one:[2]

$$w_{i\ell} = \frac{\tilde{w}_{i\ell}}{\sum_{j \in M(\ell)} \tilde{w}_{j\ell}}, \qquad \mathrm{MDI}_\ell = \sum_{i \in M(\ell)} w_{i\ell}\,\mathrm{MDI}_i.$$

To continue the example above, *Alabama* is linked to *Creek* and *Seminole* at distance 2, and to *Chickasaw* and *Choctaw* at distance 3. The row maximum is 3, so the inverted weights are 2, 2, 1 and 1, summing to 6. The normalized weights are therefore 1/3 for each of *Creek* and *Seminole* and 1/6 for each of *Chickasaw* and *Choctaw*. Our alternative measure discards the more distant pair and averages *Creek*'s and *Seminole*'s MDIs with equal weights.

[2] The unit added above ensures that the most distant match retains strictly positive weight; without it, any match at the row maximum would drop out of the average. The weights depend only on differences in distance within a language, so what matters is how much closer one match is than another. Languages with a single direct match receive a weight of one on that match.

B.2 Data Sources and Construction

Languages: Data on languages comes from the ISO 639-3 database, which uniquely identifies 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) with a three-letter code (8,552 entries total). ISO 639-3 defines language boundaries using both linguistic criteria (e.g., intelligibility) and considerations of shared versus distinct ethnolinguistic identity among speakers. Dialects—defined as divergent varieties of a language with which they share similar lexical, phonological, or grammatical features—are generally not categorized as separate entities. Only "full languages" are eligible to receive a code, i.e., those that are used in a variety of domains and support communication across all genders and all ages. Based on this definition, avoidance languages, secret languages, ceremonial languages, play languages, urban languages, youth languages, and special-purpose languages are not within the scope of this classification (SIL International, 2023).
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 each language spoken in a given country one of 12 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-1 for the full scale. We code a language in a given country as *extinct* if it falls into one of three top categories: *nearly extinct* (only remaining speakers are members of the grandparent generation or older, with little opportunity to use the language; e.g., Dakota in Canada), *dormant* (the language serves as a heritage identity marker but lacks proficient speakers; e.g., Karuwali in Australia), or *extinct* (no longer used, with no speakers retaining associated ethnic identity; e.g., Puquina in Peru).
Source: `https://www.ethnologue.com/`

Male Dominance Index (MDI): Information on language group's gender norms comes from the male dominance index (MDI) developed by Guarnieri and Tur-Prats (2023). The MDI is constructed at the ethnic group level from ten variables in the Murdock Ethnographic Atlas (EA): matrilineality, polygyny, bride price, plough use, nuclear family structure, and dependence on fishing, gathering, non-herding animal husbandry, pastoralism, and shifting agriculture. The index is the first principal component of these variables, rescaled to the unit interval. See Guarnieri and Tur-Prats's (2023) Online Appendix for details on the construction of the MDI and the EA variables included therein. For the matching of the MDI to languages, see Section B.1.
Sources: `https://doi.org/10.7910/DVN/B3LJGQ`

Societal Dominance Index (SDI): The SDI is constructed at the ethnic group level as the first principal component of three Murdock EA variables: political complexity (v33), class stratification (v66), and type of slavery (v70). The first component loads positively

on all three variables and explains 62% of the total variance (loadings: 0.6335 for political complexity, 0.6408 for class stratification, 0.4336 for slavery), consistent with the interpretation that it captures the concentration of social and political power. We rescale the index to the unit interval. For the matching of the SDI to languages, see Section B.1.
Source: derived from the Murdock EA (variables v33, v66, v70).

Kinship Tightness Index (KTI): Following Enke (2019), the KTI is constructed at the ethnic group level from four Murdock EA variables capturing the strength of extended family networks: extended family structure (v8), postmarital residence (v11), unilineal descent (v43), and localized clans (v15). Following Enke (2019), these variables are first dichotomized; the index is the first principal component of the dichotomized variables, rescaled to the unit interval. We use the PCA version to maximize comparability with the MDI and SDI. For the matching of the KTI to languages, see Section B.1.
Source: `https://doi.org/10.7910/DVN/JX1OIU`

Language Location: 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 languages' 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 before 1500 comes from combining 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 (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 in the original data. 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) (1931-1960). 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 (categories 1, 2, or 3).
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 BCE to the end of 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.
Source: `https://nathannunn.sites.olt.ubc.ca/files/2022/02/ethnographic_atlas_final.dta_.zip`

Class Differentiation: Data on languages' pre-colonial degree and type of class differentiation, excluding purely political and religious statuses (ranging from 0 to 4, from absence of significant class distinctions among freemen to complex stratification into social classes correlated in large measure with extensive differentiation of occupational statuses) comes from the Murdock EA, variable v66. 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`

Slavery: Data on languages' pre-colonial forms and prevalence of slave status, treated independently of both class and caste status (ranging from 0 to 3, from absence or near absence of slavery to hereditary slavery present and of at least modest social significance) comes from the Murdock EA, variable v70. 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`

European Colonizer's Identity and Colonizer's MDI: 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 the MDI and other ethnic characteristics of its corresponding EA group. Whenever a linguistic group was colonized by multiple colonial powers—which is the case for 17% of languages in our estimation sample—we take the average MDI of all the colonizers of a given country.
Source: `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. The male bias captures the rate at which men are depicted as more dominant, less submissive, less engaged in domestic affairs, and more physically active relative to women in a given oral tradition. For a description of the procedure we adopted to match the Murdock EA to languages, please refer to Section B.1.
Source: `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 3.4.
Source: https://archaeology.datastations.nl/dataset.xhtml?persistentId=doi:10.17026/DANS-25G-GEZ3&version=2.0

Ancestral Geographic Distribution of Ethnolinguistic Groups: Information on the spatial distribution of ethnolinguistic groups in ancestral times comes from the Murdock map, which is available only for the African continent. It is based on primary sources and is meant to represent the location of ethnic homelands during the period from 1890–1910.
Source: https://nathannunn.sites.olt.ubc.ca/files/2022/02/DATA-The-Slave-Trade-murdock_shapefile.zip

Contemporary Geographic Distribution of Ethnolinguistic Groups: Information on the spatial distribution of ethnolinguistic groups in the present day comes from the 21st edition of the Ethnologue map, which provides a shapefile indicating the location of a specific language/dialect as of the date of publication (2018).
Source: https://www.ethnologue.com/

Expansion Measure: For each Murdock group, we compute the difference between the number of grid cells in which the associated Ethnologue language is spoken today (using the Ethnologue map) and the number of cells covered by the language in its ancestral homeland (using the Murdock map). Because D-PLACE links Ethnologue languages to multiple Murdock groups, we aggregate coverage across all associated ancestral homelands when computing ancestral language use, to ensure comparability between ancestral and contemporary extents. We then define expansion as this difference, set negative values to zero, and take the log of one plus the resulting value.
Source: Ethnologue map: https://www.ethnologue.com/
Murdock map: https://nathannunn.sites.olt.ubc.ca/files/2022/02/DATA-The-Slave-Trade-murdock_shapefile.zip
Grid cells: https://grid.prio.org/#/download

Frequency of Military-related Motifs: We measure military capacity at the Murdock-group level following Michalopoulos and Xue (2021), using the frequency of military-, weapon-, and soldier-related motifs in each group's oral traditions. Because these Folklore tags overlap, we take the maximum count across the three categories for each group to avoid double counting, and define military capacity as the log of one plus this value.
Source: https://doi.org/10.7910/DVN/IXOHKB

Frequency of Fertility-related Motifs: We proxy for fertility at the Murdock-group level following Michalopoulos and Xue (2021), using the number of fertility-related motifs in each group's oral traditions. We take the log of one plus this number.
Source: https://doi.org/10.7910/DVN/IXOHKB

C Appendix: Historical Narratives on the Colonial Transformation of Indigenous Gender Arrangements

The following narratives, drawn from historical and anthropological scholarship, document the ways in which European colonization disrupted the gender arrangements of Indigenous societies. The cases span North and South America, with one example from colonial Nigeria to illustrate that similar dynamics extended beyond a single continent or colonizer. They are organized in two parts. Part I presents accounts of women's roles and status in pre-colonial Indigenous societies—the social structure that colonizers encountered. Part II documents the specific mechanisms through which colonial contact dismantled those arrangements, undermining the foundations on which those cultures rested.

Part I: Women's Roles in Pre-Colonial Indigenous Societies

A traditional Cheyenne saying, cited by Almeida (1997, p. 768), encapsulates the centrality of women to the survival of their people: "A people are not defeated until the hearts of its women are on the ground." This principle was embedded in social structure itself across Plains societies. As Spack (1997, p. 35) observes, "among most Plains people, power and cultural knowledge were accumulated by and dispensed through females." More broadly, Portman and Herring (2001) document that Native American Indian women "were integral to the economic and social survival of their nations and tribes," that they were "responsible for keeping the oral traditions alive and passing them on to the future generations," and that in many Native American cultures, "women are viewed as extensions of the Spirit Mother and as keys to the continuation of their peoples."

This was equally true among the Iroquois of northeastern North America. Prior to the upheaval of the American Revolution, both Iroquois men and women understood agriculture to be women's work, carried out alongside child-rearing and household responsibilities. Men's roles traditionally centered on hunting, warfare, diplomacy, and trade. Iroquois society was matrilineal and organized around a mix of matrilocal and patrilocal residence patterns (Holly, 1990). At the core of this social order lay, in Holly's words, "a profound complementarity of or balance between male and female roles as co-important and co-fundamental economically, politically, spiritually, and biologically" (Holly, 1990).

Part II: Colonial Mechanisms of Transformation

The colonial encounter systematically dismantled these gender arrangements. Because it conflicted with their own cultural belief structures, as described by Jaimes and Halsey (1992), European colonizers worked to diminish the status of Native American Indian women within their own nations (Portman and Herring, 2001). As a consequence, in the words of Green (1992) as cited in Portman and Herring (2001, p. 195), "the economic, political, and social status of Native American Indian women suffered immeasurably."

The dynamic is captured with particular clarity by Devens (1986) in her study of New

France. The French missionaries who encountered Indigenous peoples found their balanced gender arrangements "baffled and sometimes horrified them"—irreconcilable with the French social order, in which monogamous and permanent marriage, male authority, and the control of women's sexuality and activities were the foundations of social organization. As Devens (1986) explains, the priests understood that polygamy, divorce, and sexual freedom had to be eliminated if Christian communities were to take hold. The result was a direct collision: "interaction with Europeans through missions and trade disrupted this balance," as "European concepts of social relations, combined with religious and economic programs to colonize New France, undermined the equilibrium of native communities" (Devens, 1986). Devens (1986) further documents the specific mechanisms through which this was achieved: "Priests and colonial officials deliberately and persistently stressed the importance of male authority in the community and family, and women's obligation to obedience," while the settled life and nuclear family they promoted "encouraged the breakdown of the flexible, multifamily units which were apparently based on matrilocal principles. This further weakened women's position in the community" (Devens, 1986).

Similar dynamics unfolded among the Cherokee in the southeastern United States. Perdue (1998) documents how European colonizers, in their efforts to "civilize" the Cherokee people, replaced women with men as farmers, undermined the principle of matrilineal descent, outlawed polygamy, and replaced the extended family or kin-based network with the nuclear family structure. Allen (1992) captures the political dimension of the same transformation: "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."

The Iroquois case offers a detailed illustration of how this unfolded in practice. In the aftermath of the Revolutionary War, the Iroquois faced not only military defeat but a profound restructuring of their gender order. Into this crisis arrived Quaker missionaries, who settled on the Seneca reservation in 1798 and set about replacing women's traditional role as farmers with a male-centered agriculture, cattle-raising, and the enclosure of farmlands. For some time they "were met with great resistance from a number of women," for "farming had been considered by both sexes to be women's work and it had been done with wooden hoes; the use of iron plows pulled by draft horses driven by men was unknown" (Holly, 1990). Beyond agriculture, the Quaker ideal demanded a wholesale restructuring of domestic life. The reforms they promoted required "the abdication of [women's] traditional role as the agriculturalists who provided the economic backbone of the culture, and acceptance of the new headship of the family by the husband" (Holly, 1990). Into this upheaval stepped Handsome Lake whose visions endorsed and accelerated the changes the Quakers had set in motion. As Allen (1992) documents, Handsome Lake "encouraged the shift from woman-centered society to patriarchal society." Though never complete, the shift proved consequential: under the Code of Handsome Lake—a tribal adaptation of Euro-American norms—the Longhouse (the institutional heart of clan life) declined in importance, and Iroquois women were gradually subordinated to Christian patriarchal structures (Allen, 1992).

The Jesuit missionary program among the Montagnais-Naskapi of northeastern Canada followed a parallel logic. As Jaimes and Halsey (1992) document, the Montagnais-Naskapi's dependence on colonizers compelled them to accept more of what was presented as European "morality." The Jesuits "imposed a form of monogamy in which divorce was forbidden, implemented a system of compulsory Catholic education, and refused to deal with anyone other than selected male 'representatives' of the Montagnais and Naskapi in political or economic affairs (thus deforming the Indian structure of governance beyond recognition)"(Jaimes and Halsey, 1992).

A parallel dynamic is documented among the Barí people of Colombia. Buenaventura-Posso and Brown (1980) document how the expansion of Western capitalist systems into egalitarian communal societies entails the imposition of "features such as the structure for general male dominance over females based materially on preferential access to strategic resources; private ownership of land and 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".

The same dynamics extended outside the Americas. Writing on colonial Nigeria, Oyěwùmí (1997, pp. 124–125) documents how the British systematically excluded Yoruba women from the state structures they had previously occupied, finding it "unthinkable for the colonial government to recognize female leaders among the peoples they colonized", and accomplishing "the transformation of state power to male-gender power" by "the exclusion of women from state structures."

D Appendix: Male Dominance as a Correlate of Language Survival in a Spatial Setting

In this section, we ask whether male dominance is a correlate of survival in the granular spatial dataset that we construct to conduct the analysis of Section 6. Specifically, we test whether grid cells inhabited by more male-dominant ethnic groups are more likely to have retained the ancestral language. To this end, we estimate the following specification:

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

where survival_{ig} is a 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 Section 4, constructed at the grid cell level.[3] 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, and an indicator for whether a cell is split by a country border.[4] Standard errors are clustered at the historical ethnic homeland level. Table A-27 presents the results. As the positive and significant estimate for β reveals, cells of more male-dominant ethnic groups are more likely to have retained their language, a finding consistent with what we uncover in Section 4.

Table A-27: *Male Dominance and Language Survival in Space*

	Dep. var: language survival (0-1)			
	(1)	(2)	(3)	(4)
MDI	0.42***	0.42***	0.60***	0.31*
	(0.152)	(0.159)	(0.172)	(0.164)
Mean dep. var.	0.492	0.492	0.493	0.489
Observations	15120	15120	13705	13526
Clusters	742	742	724	712
Adjusted R^2	0.017	0.019	0.051	0.074
Group size		✓	✓	✓
Distance from neighbor		✓	✓	✓
Cell split by country border		✓	✓	✓
Geography and climate			✓	✓
Other controls				✓

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 neighbor" captures the distance, in degrees, between each grid cell and the closest neighboring 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 latitude, 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.

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

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