CSEF

Centre for Studies in Economics and Finance

WORKING PAPER NO. 774

Ancestral Cultural Traits, Colonialism, and its Legacy

Marcello D'Amato and Francesco Flaviano Russo

March 2026

University of Naples Federico II

University of Salerno

Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS AND STATISTICS – UNIVERSITY OF NAPLES FEDERICO II
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 774

Ancestral Cultural Traits, Colonialism, and its Legacy

Marcello D'Amato* **and Francesco Flaviano Russo†**

Abstract

We explore whether and how the similarity of pre-existing cultural traits between ethnic groups in the former colonies and colonizers contributes to explain the legacies of colonization. We find higher levels of income per capita, and a lower probability of a "Reversal of Fortunes", in the territories where the local population had more similar oral traditions to the colonizers and where the dispersion of this folklore similarity was smaller. Exploring the mechanisms, we find that more oral tradition similarity, and less dispersion, are associated with more similar (de iure) constitutions established at independence, a higher frequency of a direct colonial rule, more conversions to Christianity and better education.

JEL Classification: J15, Z10

Keywords: Colonial Relationship; Culture; Orality; Folklore Narratives; Historical Development.

Acknowledgements: We would like to thank Graziella Bertocchi, Alberto Bisin, Giuseppe De Feo, Sergio Marotta, Massimo Morelli, Lucio Scandizzo, Marco Stimolo, Alessandro Tarozzi for useful conversations and comments. Albachiara Abate provided excellent research assistance. Financial Assistance by the Italian Ministry of the University and Research is gratefully acknowledged, Project: PRIN – CUP D53D23017970001. All remaining errors are ours.

* University of Naples Federico II and CSEF, University Suor Orsola Benincasa. E-mail: marcello.damato@unisob.na

† University of Naples Federico II and CSEF. E-mail: francescoflaviano.russo@unina.it *(Corresponding author)*

Quand les hommes ne se comprennent [...] C'est alors que la politique se tait et que la violence parle.

Albert Camus, Chroniques Algériennes (1939–1958).

"You see, colonialism is not a simple thing. It's a system built on interests."

William Walker (Marlon Brando), Queimada, 1969.

1 Introduction

Colonialism reshaped political and institutional arrangements, with long-run effects on socio-economic outcomes often stretching beyond independence (Fenske et al. 2025). But Colonialism itself was not a uniform administrative practice/rule identically applied to colonies across space and across colonizers[1], so its long-run effects clearly depend on how it was implemented. La Porta et al. (1999 and 2008), for instance, view colonization as an exogenous transfer of the colonizers legal traditions, with different effects due to the consequences of those traditions for property right protection and market regulation. Acemoglu, Johnson, and Robinson (2001, 2002), in their seminal contributions, also frame colonization as a shock to institutions, but determined by the local conditions in the colony: exogenous variation in settler mortality determined the type of institutions established, either oriented to extraction from the local population in case of a less-favorable environment to migration, or to property right protection in case of lower settlers mortality, thereby influencing long-run economic outcomes.

In this paper we take a different perspective, focusing on the cultural identity of the populations in the colonies. Colonial officers and missionaries, in fact, repeatedly confronted what

[1] Just to name few examples, British West Africa alone encompassed highly centralized emirates, decentralized acephalous societies, coastal trading polities, and ethno-linguistically fragmented regions. French Africa was similarly diverse, ranging from Islamic sultanates to animist chiefdoms to relatively egalitarian forest societies. The Iberian empires governed everything from Andean empires to Caribbean chiefdoms.

they described as the problem of governing culturally distant populations, and their accounts reveal a striking consensus: the feasibility, cost, and stability of different modes of colonial government depended heavily on the cultural characteristics of the societies in the colonies[2]. Historical and fieldwork in ethnography and anthropology (Tennebaum 1943, Fried 1952, Service, 1955), align with this view, deeming crucial, among the preexisting conditions in the colonies, the cultural similarities with the colonizer. Service (1955), in particular, in a study of the Indian-European Relations in Colonial Latin America, highlights how the "European colonists were faced with truly remarkable cultural and demographic differences among the Indians they encountered in the New World", noting also how the policies and institutions established by the colonists "varied greatly from one region to another as response to the problem posed by the extensive differences in the native cultures which they encountered". In his anthropological perspective, Service highlights an important difference in the transformative impact of the colonial experience on those population, and territories, that were forced to adjust to the new reality of the colonial rule: "The more alike the conquerors and conquered, the more simple and easy the adjustment will be, other things being equal." (Service, 1955, p. 416).

In general, pre-existing cultural similarities between the colonizer and colonized influenced the emergence of modern institutions in former colonies through their impact on the legitimization, and effectiveness, of the enforcement of both public law (organization of the administration) and private law in labor, societal and property relations. Paraphrasing Max Weber's famous definition of a modern[3] state, a colonial relationship can be considered as the establishment of a legitimate monopoly for the use of force on a community in a foreign ter-

[2]Lord Hailey (1938) famously observed that "the greater the linguistic and cultural distance, the heavier the administrative burden becomes." Similarly, the 1925 Advisory Committee on Native Education in British Tropical Africa insisted, in its well-known white paper, that European institutions "cannot be imposed upon peoples whose mentality, aptitudes, occupations, and traditions are fundamentally different". Lyautey, the architect of the French protectorate in Morocco, similarly emphasized that "administration cannot succeed if it ignores the beliefs and customs that form the foundation of the country (Lyautey 1927). Jacques Berque, one of the leading French sociologists of North Africa, later concluded that "the colonial administration failed whenever it did not comprehend the internal logic of the societies it sought to transform" (Berque 1962). Also José de Acosta (1590) described the challenges of colonial administration as rooted in "the difference of languages, customs, and ways of government, entirely other and unfamiliar to us." In short, cultural distance from the colonies was historically salient and, most importantly, its recognition was not merely rhetorical—it shaped administrative practice, resulting in marked differences even within the same colonial empires that often spanned through linguistically, religiously, and ethnically heterogeneous polities

[3]From the 1908 lecture: "Politics as Vocation". Note that term "modern" in the definition strictly applies back to institutions like European monarchies or colonial empires.

ritory, to be used more or less arbitrarily or coercively to regulate interactions among private citizens or between the citizens and the government.

The main hypothesis, then, is that the colonial rule was implemented less forcefully and coercively, and institutional transfer was easier, where the pre-existing cultural similarity between colonizer and colonized societies was higher, leading to more functional political institutions in the new order, to smoother societal transformation in the so-called modernization process, and to better long-run socioeconomic outcomes. Moreover, where the colonizer confronted more heterogeneous local populations in terms of cultural similarity, institutional transfer was arguably more complicated, and, more generally, the administration of the colonial rule presented more challenges, resulting in worse long-term outcomes. Our goal is to test this hypothesis.

To measure the cultural similarity between the colonizers and the populations in the colonized territories, at the time of colonization, we propose to exploit the information contained in the Folklore and Mythology database by Yury Berezkin (2015), which features a rich set of motifs (narratives, myths and symbols) present in the oral traditions of more than 900 ethnic groups and cultures around the World. In particular, we use the cosine similarity measure of folklore developed by D'Amato and Russo (2025 I) which, starting from the elementary units in the catalog, text strings, uses Latent Semantic Analysis to evaluate their similarity, producing an aggregate index for the extent to which the oral traditions of two ethnic groups, embedded in those motifs, overlap. Importantly, we highlight that the notion of culture underlying our empirical exercise is not that of a preeminently deep, highly persistent, trait that still affects culture today as such. It should be rather interpreted as a measure of the similarity in the configuration of cultural traits (Acemoglu and Johnson 2025), that pre-existed colonization and that, after the encounter may have shaped the social, institutional and political reorganization of the population in the colonies, with possibly persistent effects.

We construct two separate samples of territories that, at some point in history, were under colonial rule: one is at the country-level and, similar to Acemoglu et al. (2001), with unit of observation equal to the former colonies. The second is disaggregated at the ethnic group level, in the spirit of Michalopoulos and Papaioannu (2013), with unit of observation equal to the 1x1 degrees cell around the centroid, or ancestral location, of the ethnic groups in

the former colonies. In the former sample, we compute both a median over the distribution of all ethnic level folklore similarity measure between the colonizer and all ethnic groups with centroid within the colony, and a standard deviation of folklore similarities to measure also the dispersion of the similarity in cultural traits between the colonized groups and the colonizer. In both cases, we focus on European maritime colonization (Landes 1998), namely overseas empires built by England, Spain, France, Portugal, Germany, Belgium, Italy and Netherlands. Since we focus on the pre-colonial ethnic groups locations, we abstract from the effect of modern migrations.

For identification, we exploit the quasi-random assignment of a colonizer to the population in the territories of former colonies, an assumption which we motivate extensively, using both historical accounts and several others pieces of empirical evidence. In a nutshell: the colonial relationship was culturally blind, in the sense that the historical process of colonization was mainly driven by factors that can be considered as independent of the ancient cultural distance between the colonizer and the colonized. Indeed, detailed ethnographic studies were made available only after colonization, mainly as a consequence of it, so the characteristics of the local population were largely ignored at the time of colonization. An ignorance which is also reflected in how European powers agreed on a division of the non-European territories, starting from the Tordesillas treaty of 1494, according to which Spain and Portugal used an arbitrary demarcation line (based on meridians that were determined with great uncertainty given the tools available at the time) to assign newly discovered territories and those yet to be discovered. A similar logic underpinned the Berlin conference that resulted in an allocation of African territories along arbitrary straight borders.

Furthermore, to account for the possibility of endogeneity stemming from oral tradition exchanged during the colonial period especially in settlers colonies, as in the case of the diffusion of other cultural traits such as religion, we compute the folklore similarity measure by extracting a subset of "deep" cosmological motifs from the catalog, those that are generally believed to be the oldest by ethnographers (Berezkin 2015), pre-dating, among others, modern monotheistic religions and post-1500 migrations, therefore making our cultural similarity measure predetermined with respect to the colonization era.

Our main empirical test consists in regressing modern real income measure on folklore

similarity and controls. We use the GDP per capita measured in the '90s, in the country-level sample, and the geographically disaggregated GCP (Gross Cell Product) measure of real income per capita of Nordhaus and Chen (Nordhaus 2006), in the sample at the ethnic group level. Among the baseline set of controls we include: the geodesic distance from the colonizer, geo-climatic factors including settlers' mortality, pre-colonization socio-economic conditions and, importantly, colonizer fixed effects. In the cell-level dataset, we also include country fixed effects to soak-up unobserved country-level confounders. The cell-level database is particularly apt for this test because it is consistent with the identification hypothesis, given that colonization entailed a random allocation of a colonial power and rule to ethnic groups, not to (at the time non-existing) countries.

In both the country-level and in the ethnic-level samples, we find that folklore similarity between the colonized and their former colonizer is positive and significantly associated with modern income per capita. The result is considerably robust, and it holds conditionally on linguistic and religious distance, on state antiquity at the time of colonization, on migratory distance from East Africa, on the duration of the colonization period, and on the type of colony, that is on the difference between settler colonies, with more intensive migration from the colonizer, and extraction colonies. In the country-level database, we also find that the measure of variability of folklore similarity is negatively and statistically significantly associated with modern income per capita. Moreover, we find evidence that folklore similarity between the ex-colony and the former colonizer is associated with more political stability, less frequent episodes of civil conflict, better regulatory quality, better control of corruption, and better government efficiency, as well as more protection against expropriation.

We also find evidence of a negative association between folklore similarity and the probability of a "Reversal of Fortunes" (Ferguson 2003), that is there is a lower probability that relatively richer societies in the pre-colonial period became relatively poorer after colonization in case of more folklore similarity. Moreover, we find that the probability of such reversal is higher in case of higher variability of folklore similarity. For these results, we develop a new test based on a rank comparison, namely ranking all former colonies according to the pre-colonial population density in 1500, then rank them again using income per capita in 1995, code a dummy equal to one in case of at least one position lost in the ranking, and finally

regress it on folklore similarity and controls. These results are also robust to the introduction of settlers colony dummy, which is itself statistically significant, in line with Acemoglu et al. (2002).

We then explore three main mechanisms linking cultural similarity in the colonial relationship to modern outcomes. Indeed, recent contributions, surveyed in Nunn (2014) and, more recently, in Fenske (2025), emphasize that colonial legacies operated through multiple, possibly interacting, channels, reorganizing political institutions, social relations, and belief systems in ways that seem to depend not only on local geographic conditions, but also on historically rooted social and cultural configurations preexisting the colonial relationship (see also Nunn 2012 and Lowes et al. 2017).

First, we explore whether cultural similarities in the colonial relationship affected political institutions, using extensive data from the "Comparative Constitutions Project", which codes several characteristics of the constitutions, at yearly frequency, for nearly all World's countries. We build a synthetic index of constitution similarity between the ex-colony and the ex-colonizer at the time of independence, leveraging all the available information in that data base. We find evidence of more similar constitutions in case of higher median folklore similarity, and of less similar constitutions in case of more variability of folklore similarity in the colonized groups. This result holds conditionally on the identity of the colonizer, on the characteristics of the colony, on its administrative rule, and on the pre-colonial state antiquity. These findings support the hypothesis of an easier institutional transfer, or higher permeability to the colonizer formal institutions, in case of higher similarity of pre-existing cultural traits, which is also consistent with the theoretical frame proposed in Acemoglu and Johnson (2025) for the analysis of the links between culture and institutions in historical development. More generally, this result can inform the "culture versus institutions" debate (Tabellini 2012; Nunn 2014; Acemoglu and Johnson 2025).

Second, we explore the relationship between cultural similarity and the administrative rule in the colony, using the indicator of the "Form of Political Domination" from the Colonial Transformation Database. The main distinction is between a direct colonial rule, featuring a nearly total administrative control of the colony for the colonizer and no role of the local population, and an indirect colonial rule characterized by a much weaker central control and

by local administrative structures mostly in place. We find evidence that, conditional on the identity of the colonizer and on several other covariates, higher median folklore similarities between the ex-colony and the former colonizer is associated with a higher probability of a direct colonial rule. We also find that more variability of folklore similarity within the colony predicts a lower probability of a direct rule. This evidence is consistent with the recent consensus emerged both in the economic history and in the political science literature, highlighting how direct colonial rule tended to embed stronger, more uniform administrative institutions, while indirect rule left weaker and more fragmented state structures (Lange 2004 and 2009; Mamdani 1996; Iyer 2010; Michalopoulos and Papaioannu 2013; Bolt and Gardner 2020), with effects on post-independence development paths.

Third, we find evidence that folklore similarities between the colonizer and the colonized fostered conversions to Christianity and education in the colonies, most likely through missions. This result is in line with recent works (see Fenske's 2025, among others) showing that the colonial experience affected broader trajectories of social and cultural transformation, systems of beliefs and social structures of former colonies.

Our main contribution here is to study systematically how pre-existing cultural traits of the indigenous population shaped the colonization process, the implementation of the colonial rule and its effects in terms of institutional transfer and of long-term socio-economic outcomes. Thus we complement the existing explanations of the differential effects of colonization based on settlement feasibility and disease environments (Acemoglu et al. 2001), on factor endowments and inequality (Engerman and Sokoloff 2000, among others), on the pre-colonial political complexity in the colonized society (Michalopoulos and Papaioannou 2013, among others), on the transplantation of legal traditions (La Porta et al. 2008). We also contribute to the literature that documents how the colonial rule affected development through coercion, religion, education, and social organization (Nunn 2008, 2012).

The rest of the paper is organized as follows. Section 2 discusses the related literature in greater details. Section 3 explains extensively the construction of the folklore similarity measures and their determinants. Section 4 briefly summarizes the motivating evidence. Section 5 describes the samples and the estimated models. Section 6 discusses the identification assumptions. The main results and their robustness are discussed in section 7. Section 8

summarizes additional empirical results. Section 9 summarizes the result on the reversal of fortunes. Section 10 discusses the empirical evidence on the mechanisms behind the main results. Section 11 concludes.

2 Related Literature

There is a vast interdisciplinary literature on the long-term effects of colonization, spanning history, economics, sociology, and political science. Our aim is not to provide an exhaustive account, for which we refer to the surveys by Nunn (2014), Guardado (2022), Morales Arilla et al. (2022), and Fenske et al. (2025). Our goal here is only to review the most closely related works in order to better highlight our contribution.

The role of colonization in explaining the divergence between North and South America is classically associated with the work of North (1998), who highlights institutional differences between British and Spanish colonies. La Porta et al. (1997, 1998) argued that an essential aspect of the colonial legacy is that different legal traditions were transplanted according to identity of the colonizer. Bertocchi and Canova (2002) showed that long-run outcomes differ according to the identity of the colonizer and the political status of the territory. Other papers studied differences in the enforcement of the colonial rule between colonizers, mostly focusing on differences between Britain and France in Africa, such as Lee and Schultz (2012), Padró I Miquel and Yared (2012) and Letsa and Wilfahrt (2020). Differently from these contributions, we focus on within colonizer variation, that is conditioning the anlysis on the identity of the colonizer.

Acemoglu, Johnson, and Robinson - AJR in what follows- (2001, 2002), in their seminal contributions, argue that settler mortality, as defined by the disease environment faced by the colonists, shaped the type of institutions transplanted to the colonies. Similarly, Sokoloff and Engerman (2000), documenting heterogeneous growth patterns within both British and Spanish ex-colonies, argued that factor endowments and initial conditions were more important than the colonizer's identity per se.

Our research follows the logic of AJR (2001) by identifying pre-existing local conditions as primary determinants of institutional legacy. We complement their focus on biogeographical

constraints (settler mortality) by exploiting variation in the pre-existing cultural environment of the colonies. Moving beyond property rights protection as the key institutional channel, and drawing on the framework of Acemoglu and Johnson (2025), we interpret cultural similarity as an indicator of the social background of the colonial encounter that influenced the coercive intensity of colonial rule. We find that closer cultural configurations facilitated a less conflictual establishment of the monopoly on force, favoring the implementation of direct administrative rule, the assimilation of formal constitutions at independence, and deep social transformations in education and religious beliefs.

Previous studies found evidence of a "Reversals of fortunes" (Acemoglu et al. 2002, Lange et al. 2006, Nunn 2008 and Chanda et al. 2014, among others), according to which richer societies in pre-colonial times ended up with lower values of income per capita today. Our contribution on this literature is both to develop a new test for reversal, based on rank comparisons in socio-economic indicators, and to provide evidence on the importance of pre-colonial cultural similarities in affecting the probability of reversal. Importantly, the persistent legacy of cultural similarity in the colonial relationship should not be interpreted as due to hard wired cultural traits that have persisted. By highlighting the role of pre-existing cultural characteristics in the reversal of fortunes, we show how cultural similarities may have affected historical development paths in former colonies.

Several works already highlighted the effect of cultural similarity with the colonizer in shaping the colonial relationship through specific colonial rules, within the same empire. Early examples include the historical accounts by Lugard (1922), Hailey (1938), Lyautey (1927), Delafosse (1912), Service (1955), and Perham (1962), who share the idea that linguistic, religious, and institutional distance from the colonizer often made a direct administrative rule in the colony rule infeasible, prompting reliance on indirect rule through customary authorities. Recent contributions in political-economy, such as Gerring et al. (2011) show that indirect rule was more prevalent where colonizers faced greater cultural and administrative unfamiliarity. Among others, Bolt and Gardner (2020) and Hariri (2012) emphasize the role of pre-colonial political fragmentation and ethnic heterogeneity in shaping colonial governance strategies. In general, this literature supports the notion that cultural and ethnic structures mattered not only for post-colonial development, but also for the administrative technologies deployed by

the colonial powers (on this point we provide further information in the Appendix). Our contribution builds on these insights, providing systematic evidence that cultural similarity shaped the extent of institutional transfer within empires and that this mechanism is distinct from ethnic fragmentation, pre-colonial political complexity, and the settler-mortality channel.

As for the long-term effects of different colonial rules, Lange (2004, 2009), among others, within the context of the British Empire, shows how colonies with greater British institutional penetration, that is with a more direct administration rule, developed stronger rule-of-law institutions, higher bureaucratic capacity, and, ultimately, significantly higher contemporary GDP per capita. In contrast, regions governed through indirect rule— where colonial authorities tended to delegate power to traditional chiefs and customary courts— exhibit weaker legal institutions and poorer economic performance today. Iyer (2010) similarly studies the effect of different colonial rules on long-term outcomes in India. In Africa, Mamdani (1996) argues that an indirect rule produced "decentralized despotism": colonial powers empowered unelected local elites with discretionary authority, creating rigid ethnic hierarchies and legal pluralism that undermined the emergence of modern citizenship and state institutions. According to this view colonial regime based on indirect rules ended up to freeze traditional power structures and socio- economic equilibria in the colonies, whereas these power structures had been eroded in the course of the XVI-XVII century in Europe. These institutional distortions persisted after independence, contributing to lower development and higher political instability. In our analysis we find that that more similar pre-existing cultural traits between the colonizer and the colonized affects the probability of observing a direct rule in the colonial relationship.

Many contributions have studied the effects of the transplantation of formal legal traditions and political institutions, decades after independence (see Nunn 2014 and Fenske 2025 for a survey). Our exploration of the relationship between the similarity of pre-colonial cultural traits and those in the formal political constitutions of former colonies at independence provides corroborating evidence that formal political institutions represent an important channel to understanding the legacy of colonialism. This evidence, we believe, is relevant for the discussion on the relation between culture and institutions as in Acemoglu and Johnson (2025), Tabellini (2012), Nunn (2014).

Our work is related to Michalopoulos and Papaioannou (2013), who find a strong asso-

ciation between pre-colonial ethnic political centralization and regional development at the ethnic group level in Africa. Whereas they focus on pre-existing institutional conditions of the colonial relationship, our focus is on pre-existing cultural similarity in oral traditions, which we also explore, following their contribution, at the ethnic group level. Michalopoulos and Papaioannou (2016) document how colonial boundaries, largely drawn at the Berlin Conference of 1884–85, frequently split ethnic groups and were drawn with little regard for indigenous settlement patterns, with long-lasting consequences for political stability and income. Their key insight is that borders were, to a large extent, exogenous to local pre-colonial characteristics. This is also part of our identification assumptions.

The paper is also related to the literature on the role of religion and missions in historical development, to the extent that pre-colonial cultural similarities affected the diffusion of alternative system of beliefs in the colonized society. Nunn (2010) studies the long-term effects of exposure to Christianity through missionary activity during the colonial period on contemporary beliefs and values in Africa. We offer new evidence that the diffusion of Christianity was facilitated by pre-colonial oral tradition similarities.

Finally, and more generally, this paper contributes to the literature on folklore in economics and other social sciences. Michalopoulos and Xue (2021) introduced the Berezkin folklore catalogue to economics and showed that folkloric content is associated with contemporary attitudes and economic outcomes. Galor et al. (2023) use folklore-based measures of cultural diversity to study how prehistoric migration patterns shaped contemporary cultural diversity and inequality. D'Amato and Russo (2025 I) show that similarities in folklore across groups and nations help explain income differences and, in D'Amato and Russo (2025 II), the incidence of interstate conflict.

3 Folklore-Based Cultural Similarity

In this section we provide a detailed description of the two measures of folklore similarity used throughout the analysis (section 3.1) and the main arguments for why they can be interpreted as a measure of the similarity of the cultural configuration between the two partners in a colonial relationship (section 3.2).

3.1 Measuring Folklore Similarity

Our benchmark indicator of cultural similarity between ethnic groups or cultures is the cosine measure of folklore similarity developed by D'Amato and Russo (2025 I), to which we refer for further details on its construction. In brief, it measures the overlap of the oral traditions of ethnic groups, starting from the Berezkin folklore catalog (Berezkin 2015) that records the presence of folklore motifs in the oral tradition of several ethnic groups and cultures. Motifs consist of narrative episodes, images, plots and characters featured in the traditional stories. Since the motif dataset features text strings, the first step of the computation entails the use of Latent Semantic Analysis, a lexicostatistics algorithm, to evaluate the textual similarity of the motifs. This allows, in particular, the computation a cosine similarity between them, that can be understood as a correlation coefficient for text strings. Folklore similarity between two ethnic groups i and j included in the Berezkin catalog is then computed as follows:

$$\lambda_{ij} = \frac{1}{|M^i \cup M^j|} \sum_{a=1}^{A_i} \max_{m_b} [C(m_a, m_b)] \tag{1}$$

where M^i is the set which includes the A_i motifs m in the oral tradition of group i, M^j is the set which includes the A_j motifs m in the oral tradition of group j, the vertical bars indicate the cardinality of the union set, and the function C denotes the cosine similarity of the motifs.

The most attractive feature of this folklore similarity measure, making it more apt to evaluate folklore similarity than simpler set similarity measures, is its ability to account for imperfect cultural transmission. For instance, suppose there are two ethnic groups, one featuring the motifs A, B and C in the oral tradition, and another featuring the motifs A and D, but with the motif D being textually similar to C in terms of narrative structure, characters or others elements of the motifs. For instance, D and C could be similar stories about deception by tricksters, with a similar plot and takeaway message (trickster punished), but with different characters (fox or monkey deceiving) and details (disguises, objects etc.). In this example, using a simple set similarity measure evaluating the overlap of the sets of motifs (motif A in the above example, over a total of 4 motifs) will underestimate the true extent of oral tradition similarity, calling for a correction based on the motifs text.

There is, in principle, the possibility that colonization, because of frequent exchanges between the colonizer and the colonized, might have led to cultural similarities as a result. For instance, loanwords due to, say, commercial interactions, probability induced lower linguistic differences (Blouin and Dyer 2021). Given that the colonial experience might have affected some of the more recent motifs featured in the Berezkin folklore catalog, to guarantee that our measure of cultural similarity is predetermined, we restrict attention to the motives in the A group according to the Catalog classification, briefly described as "Sun and Moon" motifs. According to Berezkin (2015), these are the oldest ones, and some of them probably date back to 10000 BC or before, and, importantly for the identification of the empirical model, they are much less likely to have been transferred across groups, especially as an effect of the establishment of the colonial relationship.

We use the information on the ethnic groups centroids, or ancestral locations, to match them with a colonizer and with a colonized territory, and then compute two summary statistics that will be our benchmark measures of cultural similarity in the country-level database. The first, for each colonized country k, is the median across groups cosine measure of folklore similarity between the ethnic group i in the colonizer's country[4] and all the j ethnic groups with centroid within the modern borders of the colonized country. Formally, define J^k as the number of those groups with ancestral homeland in the colony, the measure is equal to:

$$\Lambda_i^k = \begin{cases} \lambda_{i[(J^k+1)/2]} & J^k \text{ odd} \\ \frac{1}{2}\left(\lambda_{i[J^k/2]} + \lambda_{i[J^k/2+1]}\right) & J^k \text{ even} \end{cases} \tag{2}$$

where the brackets contain the medians over the sample values of λ sorted in ascending order, that is $\lambda_{i[1]} \leq \lambda_{i[2]} \leq \cdots \leq \lambda_{i[J^k]}$. A high value of Λ indicates that the ethnic groups in the colonized territory feature oral traditions with a significant overlap with those in the colonizer.

The second measure evaluates the variability of cultural similarity between the colonizer i

[4]In the benchmark implementation, we treat colonizers as homogeneous, we use just one value of i, that of the most important ethnic group in that territory, namely, according to Berezkin classification of groups/cultures: English in England, French in France, Spanish in Spain, continental Italians in Italy, Dutch/Flemish/Frisians in Netherlands and in Belgium, Germans in Germany, and Portuguese in Portugal. We find however similar results when accounting for potential ethnic heterogeneity within the colonizer. This last result suggests that they are robust after taking into account the presence of Scots and Welsh in England, Catalans and Basques in Spain and Walloons in Belgium.

and the colonized groups j with ancestral homeland centroid in the current borders of former colony k. It is computed simply as the standard deviation over all the values of λ_{ij} for each colonial relationship:

$$S_i^k = \sqrt{\frac{1}{J^k}\sum_{j=1}^{J^k}(\lambda_{ij} - \lambda_i^k)^2} \tag{3}$$

where λ_i^k is the sample average value of λ_{ij} computed over all values of j in the current borders of the colonized country k. A high value of S indicates that the colonized territory was inhabited by highly heterogeneous ethnic groups at the time of the colonial encounter, whereas a small value indicates a culturally homogeneous population, with respect to the colonizer, although it does not say if this difference is big or small. Intuitively, Λ provides a measure of the similarity of the local configuration of cultural traits recorded in the oral traditions of people in the colony with that of the colonizer, whereas S provides a measure of the cultural diversity in country k measured at the moment of the colonial encounter. Of course, in case of a single ethnic group in the colonized territory ($J^k = 1$), $\Lambda_i^k = \lambda_{ij}$ and $S_i^k = 0$.

In the ethnic group level database, by definition, there is just one colonized ethnic group j in the area that corresponds to our observation unit, therefore we only use the value of λ_{ij} as main measure of cultural similarity since there is no standard deviation to consider. For the computation of Λ, we prefer to use a median, rather than the average, although we find fully robust results when using the average.

3.2 Determinants of Folklore Similarity

To clarify the reasons why we propose the two measures of cultural similarity Λ and S to quantify the extent of cultural diversity in the colonial relationship, we now briefly describe how this folklore similarity is determined. According to ethnologists, there are two main mechanisms leading to folklore similarity, convergent cultural processes or cognitive universals, and ancient cultural diffusion.

Convergent cultural processes, which is akin to convergent evolution in biology, entails the possibility of independently developing similar folklore motifs in the absence of a direct cultural exchange and of cultural diffusion. The main assumption is that there is an underlying

similar human psychology or universal cognitive structure that might have lead to similar oral traditions (Tylor 1871; Dundes 1980). Although Berezkin argued that this convergence channel is a rather small determinant of folklore similarity (Berezkin 2023), we think that, especially for the A group motifs which we are focusing on, this is indeed a concrete possibility. After all, simple observations of the sun and moon cycles, and of the moon shapes and shades of color, could have spurred the separate development of similar traditional stories. Convergent cultural processes can also be the result of common exogenous shocks such as similar geo-climatic conditions or similar events such as floods or earthquakes, although this is only a marginal source of similarity of the motifs related to sun and moon.

Pre-colonial cultural diffusion refers instead to the possibility of ancient cultural contacts that determined the horizontal transmission of narratives across different ethnic groups, following migration, interbreeding, and trade patterns, or as a consequence of ancient conflicts. In fact, according to Berezkin: "The replication of forms that can be borrowed from culture to culture [...] reveals the particular and, to some degree, random peculiarities of [...] cultural contacts and interactions" (Berezkin 2015, p. 34-35). This is the so-called diffusionist theory in anthropology and comparative mythology (Smith 1916; Witzel 2012; Matsumura 2019; Therani 2023; among others), which emphasizes the role of cultural contacts in the diffusion of knowledge and other cultural traits and which, in its extreme form, postulates a unique origin of the folklore motif, in sharp contrast with the possibility of convergent evolution.

We cannot distinguish, in practice, between folklore similarity stemming from cognitive universals or for ancient cultural diffusion. For instance, consider the "Figure on the Lunar Disc" motif (A32 in the Berezkin catalog classification, whose brief description is: "An object or being is seen in the moon), and which comes in different variants according to the exact shape of the object (Thuillard 2021). This motif is present in several cultures around the globe, 354 out of about 1000 coded in the catalog, for instance, among others: in the Chinese, Mongol, Miao, Sichuan, Hindi, Khmer, Sinhalese, Thai, Evenki, Tatar, and Adyghe cultures in Asia, Germans, Italians, Greeks, French and English in Europe, Aztec, Yucatec, Mixe, Quechua, Hopi and Osage in the Americas, but very few groups in Africa. The variant of this motif according to which the figure is a rabbit or a toad, most likely originated in central Asia, but that appears in a lot of south American cultures, and it is not clear, whether this is due

to diffusion, through migration, for example, via the Bering Strait or to convergent evolution, given that there is no information about the time since a motif or a story entered in the oral tradition of a group. The fact that the motif is almost absent from African cultures actually seems to exclude the possibility of a convergent evolution. Another example is the "Male sun and female moon" motif (A3 in the Berezkin catalog classification), which is present in about one third of the cultures covered by the catalog. For instance this motif is present in the oral tradition of several groups such as, among others: the Spanish, Italians and Greeks in Europe; the Marathi, Kashmiri, Yao, Sichuani and Yava in Asia; the Cree, Ojibwa, Guarani, and Aztec in America; the Western Pygmies, Akan, Ashanti, Lingala, Rwanda, Kongo, Northern Gur and Zulu in Africa. Such widespread diffusion contrast with the idea of cultural contacts being the only determinant of folklore similarity, calling into question convergent evolution.

Even if we cannot distinguish between the two main sources of folklore similarity, it is important to highlight that this does not represent a problem for our identification strategy and for the interpretation of the results. If folklore similarity was purely the consequence of cognitive universals, then it would be a clear exogenous shock, perhaps induced by similar geo-climatic conditions in the homeland environment, on which we can however condition in the analysis. If instead it was the result of cultural diffusion alone, we assume that those ancient relationships were not the basis on which the match between colonizers and colonized was determined, largely because those similarity were not known at the time of colonization and because the division of extra-European territories among European powers was independent of the identity of the populations, as we extensively discuss in section 6.

A last possible determinant of folklore similarity is vertical transmission from a common ancestor. According to the out-of-Africa hypothesis, all modern humans originated in Eastern Africa and then spread throughout the World gradually separating and forming different groups (Ashraf and Galor 2013). Assuming that traditional stories are transferred across generations, then having a recent common ancestor, that is being part of two populations that separated more recently, can be associated to a more similar oral tradition (D'Amato and Russo 2025 I). In the context of our analysis of the impact of colonization, this not a concern given the quite long separation time between populations in Europe and those in the rest of the World.

Therefore we conclude that, regardless of the reason for the emergence of the observed motif similarity, we can safely consider them predetermined at the time of the colonial encounter. In addition, in the remainder of the analysis we will use two indicators to control for common ancestry, genetic distance (Spolaore and Wacziarg 2009) and migratory distance from East Africa (Ashraf and Galor 2013).

4 Motivating Evidence

A colony is a territory subject to a foreign rule (Oxford Dictionary), after migrations, settlements, invasions or wars. In this work, we will narrow down this broad historical set, focusing mainly on those territories that were subject to "European Maritime Colonization" (Landes 1998), thus overseas empires built as a consequence of naval power and exploration with the main purpose of extraction, settlement, and control of trade routes[5]. Within this set of colonies, there are both, according to the classification by Acemoglu et al. (2001), "Settlers" colonies, characterized by a large immigrant population from the controlling country, and "Extraction" colonies, mainly used to harvest natural resources, establish plantation modes, exploit forced labor and obtain commodities to be shipped in the metropole or in the global trade network.

Following up on the seminal works by North (1998), Sokoloff and Engerman (2000), Bertocchi and Canova (1999), and Bolt and Gardner (2020), we observe that there is a considerable variability of long-run growth performances for former colonies, both in general[6] and within colonizers, that is fixing the identity of the foreign power and, therefore, the type of institutions, technologies, human capital, etc. that can be potentially transferred, and on the political sentiment or incentives in the metropolis for establishing extractive institutions

To give a sense of the magnitudes of the within-colonizer differences, consider the variability of modern GDP per capita among former colonies within the same empire, and compare it with

[5]Asian empires, on the other hand, were based on land expansion to contiguous territories using army power. We propose an extension to Asian empires in section 8.

[6]Former colonies are not found to be poorer, on average, today. Regressing income per capita in 1995 on a dummy equal to one for former colonies, in a sample of World countries, there is a negative and strongly statistically significant coefficient. However when controlling for continent fixed effects, the coefficient is not significant anymore, and adding some controls (absolute value of the latitude, share of the country in tropical and subtropical climate zones, population in 1500, neolithic transition timing) delivers the same results.

the variability of average modern GDP per capita across former empires. Across empires, the standard deviation of GDP per capita (in thousands PPP dollars) is 3.13 (interquartile - iq-range 5.8), while within former British, Spanish, French, and Portuguese colonies, respectively, 14.67 (iq range 16.98), 6.49 (iq range 7.71), 8.69 (iq range 4.36) and 20.42 (iq range 10.71).

As an example, consider, within the French colonial empire, the very different long-term trajectories of Gabon and Cameroon, in Sub-Saharian Africa. The two countries share a border, so they are characterized by similar geo-climatic conditions, both face the Atlantic Ocean, and both are heterogeneous with respect to the ethnic composition of the population, and yet Gabon has a GDP per capita that is considerably bigger. Mozambique and Angola are another example of countries within the same continent (although facing different oceans), with the same colonizer, and very different levels of income per capita, higher in Angola. As a further example, consider the case of Mauritania and Senegal, two countries that share a border and which both were within the French colonial empire, and with a higher modern level of GDP per capita in the former country.

The question, then, is what determined this variability of outcomes within empires. In what follows, we will study these differences systematically, focusing, in particular, on the role of a very peculiar pre-existing condition in the colonial relationship, the cultural similarity between the colonizer and the colonized ethnic groups.

As preliminary evidence, notice that, according to our Λ measure of median folklore similarity, the indigenous population in Gabon was more similar to the French colonizer than the indigenous population in Cameroon. Similarly, the indigenous population in Angola was much more similar to the Portuguese colonizer than the population in Mozambique. Moreover, Mauritania is characterized by just one ancestral ethnic group (Western Sahara and Mauritania culture, in the Berezkin classification), so by a zero variabilility measure S, while the modern territory of Senegal is significantly more heterogeneous, with 5 ancestral ethnic groups/cultures (Wolof, Fula, Soninke, Serer, and a composite made of Manjak, Papel, Diola and Balant). In Table 1 we report the summary statistics of folklore similarity, for all colonies and for the major colonial empires.

5 Empirical Models

We gathered data at two different aggregation levels, respectively country-level (section 5.1) and grid-cell-level (section 5.2). We discuss each of them, and the associated empirical models, in what follows.

5.1 Country-Level Dataset

The unit of observation in the country-level dataset is the ex-colony k, defined by its current borders. We estimate the following empirical model:

$$Y_{kic} = \beta_0 + \beta_\lambda \Lambda_i^k + \beta_s S_i^k + \Gamma X_{ki} + \eta_i + \theta_c + \varepsilon_{kic} \tag{4}$$

where Y_{kic} is the logarithm of the GDP per capita of the ex-colony k in continent c and whose colonizer was i, Λ_i^k is the median folklore similarity (in logs) defined in equation 2, S_i^k is the standard deviation of folklore similarity defined in equation 3, X_{ki} is a vector of control variables, η_i are colonizer fixed effects, θ_c continent fixed effects, and ε is the error term. The coefficients of interest are β_λ and β_s, where β_λ captures the effect of cultural similarity between the colonizer and the ex colony on long-run economic performance, and β_s captures the effect of the cultural heterogeneity of the local groups, at the moment of the colonial encounter. As benchmark target value for income per capita, we use its value in 1995, but we find similar results when focusing on alternative dates (1990, 2000, 2005, 2010, 2020). Since we have colonizer fixed effects in the regression, we are effectively comparing long-run performances of colonies within the same colonial empires, that is conditioning on the characteristics of the colonizer.

We consider the following control variables: the geodesic distance between the colonizer and the ex-colony, because closer colonies to Europe might have facilitated both the demographic, technological and institutional transfer, because cultural similarities might simply reflect geographic similarities, and because proximity could have influenced the choice of which country to colonize (Landes 1998); the absolute value of the latitude of the ex-colony capital, and the percentage of the country territory in tropical and subtropical climatic zones (Koeppen-Goering classification), that are associated to bio-climatic factors that might have an impact

on income per capita (Sachs 2001). Notice that both variables are correlated with settlers mortality, given that tropical diseases such as Malaria and Yellow Fever were the main causes of death. We also consider: the pre-colonial population density, as of 1500 (McEvedy and Jones 1978), which is associated with the type of institutions in the colony (extractive in case of higher population density, see Acemoglu and Johnson 2005), that predicts the possibility of reversals, as shown by Acemoglu et al. (2002), and that is related to the choice and timing of colonization (Ertan et al. 2016) and to long-run outcomes (Ang 2013). We also include continent fixed effects which capture, among others: climatic effects, axis orientation, suitability to agriculture, and other geo-climatic factors that have an impact on long-run growth; different colonization periods, as Americas were colonized much earlier than African countries. Importantly, the pre-existing conditions in the colony help capturing the effect of potential unobserved determinants of both deep folklore similarity and current income levels[7], under the assumption that their effect on development must have been evident before colonization. We also include continent fixed effects which capture, among others: climatic effects, axis orientation, suitability to agriculture, and other geo-climatic factors that have an impact on long-run growth; different colonization periods, as Americas were colonized much earlier than African countries.

We also include linguistic, religious, and genetic distance between the colonizer and the ex-colony, as discussed in D'Amato and Russo (2025 I), to make sure that we are identifying the effect of folklore similarities from other cultural traits measured in alternative indeces. For genetic distance, we use the microsatellite measure from Pemberton et al. (2013), based on neutral genes not affected by evolution, between dominant, pre-colonial, ethnic groups (Spoloare and Wacziarg 2009 and 2018). Since this can be interpreted as the separation time of two ethnic groups since they had a common ancestor (Spolaore and Wacziarg 2009), and since it is mostly influence by demic movements (Bortolini et al. 2017), including this variable allows to isolate the effect of folklore similarities determined by cultural contacts and perhaps convergent evolution, from the effects of demic movements. For linguistic distance, we use the lexical similarity measure from the Automated Similarity Judgment Program (Wichmann et al. 2022), equal to the average Levenshtein distance (number of replacements needed to

[7]For instance, folklorists believe that motifs followed the patterns of ancient demic movements (Bortolini at al. 2017).

convert one word into another) between 40 basic words/meanings (chair, water, eat, food, etc.). Importantly, given that some of these words might have been exchanged as a consequence of colonization, that is because of cultural contacts during the colonial period, we cannot assume that this distance is predetermined, unlike the folklore similarity measure that we compute using deep motifs only. For religious distance, we use the complement to one of the relative number of common nodes in the World Christian Database religious tree. As for the case of linguistic distance, there is indeed the possibility that also religious distance is not predetermined.

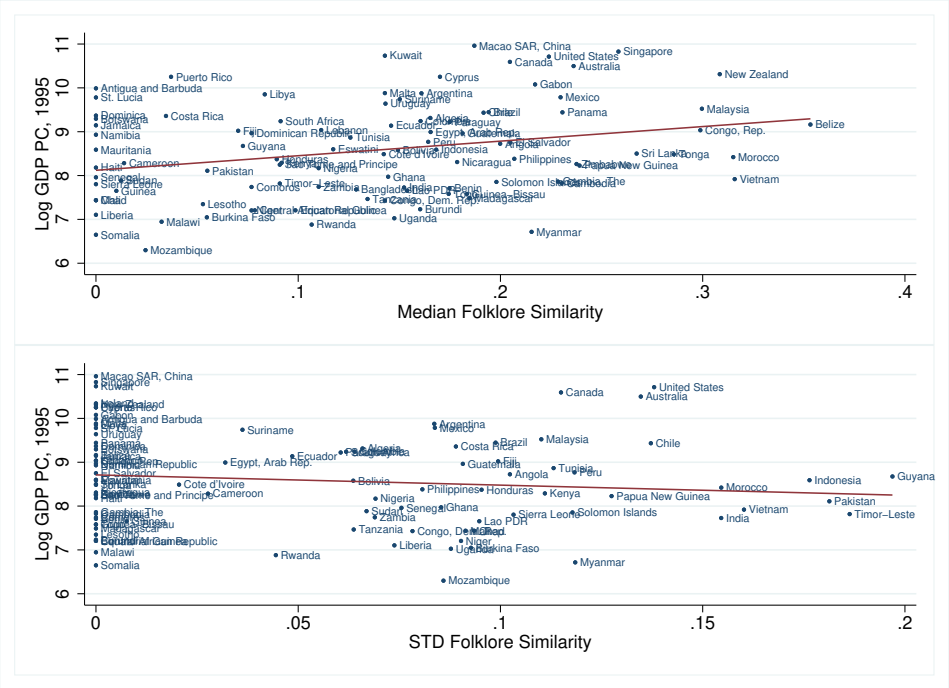
The sample is made of former colonies of the following European countries[8]: France, England, Spain, Portugal, Germany, Belgium, Netherlands, and Italy. The colonizers fixed effects will capture the main structural, unobserved, differences among them, such as the size of the dominion, the different colonization waves (earlier colonization of the Americas vs late colonizations of Africa and Asia), the specific administrative and legal traditions relevant for the establishment of the colonial rule, the main patterns of the trade network in which the specific colonial relationship was enshrined, and their overall political "power", that might have driven the (perhaps assortative) matching between the colony and the colonizer, given that colonial expansion was mostly shaped by the evolution of the balance of power between European States (see section 6 for an extensive discussion).

As preliminary evidence, in Figure 1 we plot the relationship between log income per capita in 1995 and both the median folklore similarity (upper panel) and its standard deviation (bottom panel) between the ethnic groups in the former colonies and their colonizer together with a linear fit, for all ex-colonies for which we have information.

The average value of the median folklore similarity Λ in the sample is 0.143, with a median equal to 0.133 (so the distribution is close to be symmetric), a quite high standard deviation of 0.103, an interquartile range equal to [0.037,0.194], and 20 former colonies characterized by 0 median folklore similarity, so with to to very little overlap of the oral traditions. As for the standard deviation of folklore similarity S, 47% of the former colonies in the sample are characterized by just one ancestral ethnic group. Over the whole sample, the average standard

[8]For some modern-day countries, we have overlaps of different colonizers within the borders, or exchanges of territories between colonizers. In these (few) cases, following previous literature, we assign the colonial relationship to the colonizer with the longest, or more extensive, domination in the area (see appendix for details on the assignments).

Figure 1: **Folklore Similarity to the Colonizer and Income per Capita**



Notes: Y-axis: log GDP per capita in 1995 from the World Bank Development indicators. X-axis, top panel: median folklore similarity (Λ) between the ancestral ethnic groups in the colony and their former colonizers. X-axis, bottom panel: standard deviation of folklore similarity (S) between the ancestral ethnic groups in the colony and their former colonizers. Linear regression fit. 110 observations (former colonies).

deviation of folklore similarity between the ethnic groups in the colony and their former colonizers is equal to 0.05, with median 0.024, standard deviation 0.062 and interquartile range [0,0.092].

5.2 Ethnic-Level Dataset

The second database is designed to explore our main hypothesis at the ethnic group level. Specifically, the investigation of colonialism for the territories that used to be under the control of more or less politically complex community of ancestral ethnic groups is important for at least three reasons. First, for consistency with our identification hypothesis, since the historically random allocation of colonial powers was to territories and, hence, to population groups, not to countries, which did not exist at the time.[9] Second, and relatedly, the literature

[9]The problem has been indeed closely considered before in the literature: one might question the utility of examining the effects of colonialism on territories currently governed by political units that did not actually

on the role of pre-existing conditions for the impact of the colonial rule on contemporary outcomes, such as, among others, Michalopoulos and Papaioannou (2013) has clarified that pre-existing conditions can conveniently be measured at the ethnic group level. Third, because it alleviates general aggregation bias concerns due to the bunching of group level ethnographic information at the country level.

In this database, the colonial relationship is defined at the level of the two ethnic groups involved in the relationship. Therefore the observation units are the centroids or ancestral, pre-colonial, location of the ethnic groups that, at some point over their history, fell under colonial rule. To match the ethnic group centroids to a colonizer, we use various historical maps of the colonial empires collected in the Wikimedia Commons Atlas of Colonialism[10]. In case of more than one match, we assign the colonizer that has been documented to have ruled for a longer period. This allows us to study the modern-day consequences of colonization on income per capita at a much finer geographical level.

As outcome variable, we use the Gross Cell Product (GCP) at PPP from the G-Econ database assembled by W. Nordhaus and X. Chen, which is akin to a disaggregated GDP per capita measure at a fine geographic level. This measure is available at a resolution of 1 degree latitude times 1 degree of longitude for virtually all world locations (Nordhaus 2006). We simply match each ethnic group with the closest cell of available GCP data, and we take averages over the 4 years of available data, namely 1990, 1995, 2000 and 2005. As an alternative, we also used data on night time light intensity (Henderson et al. 2012) but, since the empirical results turn out to be very similar to those obtained by using the cell product, we will not report them in what follows. The estimated empirical model is the following:

$$Y_{jic} = \beta_0 + \beta_\lambda \lambda_{ij} + \Gamma Z_{ij} + \eta_i + \theta_c + \varepsilon_{jic} \tag{5}$$

where Y_{jic} is the logarithm of the GCP per capita in the cell where the centroid of the ethnic group j is located, i is the colonizer, c is the continent where the cell is located, and λ_{ij} is the folklore similarity between the colonizer i and the ethnic group j (in logs), Z_{ij} is a set of control variables, and η_i and θ_c are, respectively, colonizer and continent fixed effects.

exist during the colonial epoch (e.g. Mahoney 2003).

[10]See the webpage for an exhaustive list of the sources used (`https://commons.wikimedia.org/wiki/Atlas_of_colonialism`).

The set of baseline control variables X is necessarily partially different in this more granular specification, due to data availability. In particular, we include: the geodesic distance between the centroids of the ethnic group and the capital city of the colonizer, the absolute value of the latitude of the colonized ethnic group centroid, the average (modern) temperature within the cell. We also include, as further controls, linguistic and religious distance between the colonizer and the colonized ethnic group. We add colonizer fixed effects in all regressions consistently with the main empirical strategy of the paper. We will present two sets of results, one with continent fixed effects, and standard error clustering at the level of the colonizer, and a second with (modern) country fixed effects, and standard error clustering at the country level. This last specification is empirically more demanding, and it relies on within country variability, meaning that we are effectively disregarding information relative to the modern countries with just one native ethnic groups within the modern borders[11]. However this specification allows to control for several country-specific, time-invariant, confounding factors.

We are able to build up a sample of 690 ethnic groups that were ever subject to a colonial rule by Europeans at some point between 1500 and the second half of the 20th century. Out of these groups of formerly colonized groups, 136 (19.7%) are located in Africa, 410 (59.4%) in the Americas, and the rest in Asia and Oceania. Breaking down by colonizer, we have 285 (41.3%) groups under British colonial rule, 153 (22.2%) under Spanish rule, 113 (16.4%) under French, 68 (9.8%) under Portuguese, and the rest for the remaining colonizers, that is Belgium, Germany, Italy and the Netherlands. This cell-level dataset is characterized by a considerable variability of income per capita within colonizers, even more so than the country-level database. In particular, the coefficient of variation of cell GCP per capita is 2.84 for cells within the British colonial empire, 3.07 within the French, 3.18 for the Spanish, 2.83 for the Portuguese, and 3.12 for the Dutch.

[11]The median number of native ethnic groups per country is 3, and 26 countries (31% of the total number of countries) have only one native ethnic group. However the ethnic groups which are the only native ones with centroid within the modern border of a country are just 3.7% of the total of ethnic groups in this disaggregated sample

6 Identification

The main identification assumption in our analysis is the random assignment of folklore similarity to the colonizers conditional on observables, which, among others, excludes the possibility of selection into colonization. In practice, various ethnic groups outside Europe, at some point in their history, were "encountered" by an arguably random stranger group from Europe, and eventually fell under its colonial rule, but the cultural identity of this foreign group was not determined by cultural characteristics of the population in the colony. In this section we briefly summarize the main arguments in favor of this assumption, which are further detailed in appendix together with some empirical tests.

Ethnographic Studies. The first important piece of evidence in favor of our identification assumption is that there was no systematic ethnographic information available at the time of colonization that might have guided the choice of where to establish a colony. The general consensus among historians and anthropologists (Asad 1973; Fabian 1983; Stocking Jr. 1987; Pratt 1992) is in fact that ethnographic studies were an effect, rather than a cause, of colonization, for several reasons. First, simply because of access, that is because contact with the indigenous population that helped assembling the material, with the help of missions and schools. Second, because knowing the indigenous population societal organization eased the administration of the colony; third because of the wish to find a "rational" justification of colonization in terms of civilization of an otherwise uncivilized population.

Historical Evidence. The second piece of evidence in favor of our identification assumption is that assignment of specific colonies, and hence the assignment of the folklore similarity between the colonizer and the colony, was randomly determined by historical events in Europe after the discovery of the New World, thereby excluding the possibility of selection.

The *Treaty of Tordesillas* is a case in point: in 1494, Spain (Castille crown) and Portugal, mediated by the Pope, signed an agreement with the intent to settle future disputes over discovered territories in the Americas, establishing a (papal) demarcation line of 100 leagues (about 320 miles) west of the Cape Verde Islands, such that all lands discovered, or to be discovered, to the west of this line would belong to Spain, and lands to the east would belong to Portugal. After a slight subsequent adjustment of the demarcation line, Portugal was

granted the easternmost part of South America, while Spain all the rest of the Americas. Moreover, Portugal secured a route around Africa, solidifying its claims to former trading posts in Angola, Mozambique, India, Malacca, and Macau. Clearly the demarcation was absolutely random and, importantly for our purposes, the agreement was not at all based on the characteristics of the discovered colonies.

Soon after, England, France and the Netherlands rejected the Tordesillas treaty's (Papal) authority, exploiting the weaknesses of Spain after the Thirty years war, and created new legal and political justifications for their own empires. According to the "Doctrine of Effective Occupation", European power had to actually settle, administer, and defend a territory to have a legitimate claim, so those newcomers in the race for colonies started establishing colonies and trading posts in "unoccupied" or poorly defended areas of the Americas, especially in the North and in the Caribbeans. This new wave of colonial expansion was therefore purely determined by the balance of power among European countries, rather than by the characteristics of the population in the colonies.

As for Africa, leading European powers competing for influence and resources- the "Scramble for Africa"- gathered in Berlin in 1884 to negotiate, in an effort to prevent major conflicts, and without representatives from African nations. The main outcome, together with a reaffirmation of the principle of Effective Occupation, was a partition of Africa into spheres of influence, colonies and protectorates, through the construction of formal boundaries. Given the very limited knowledge of the local population by Europeans at that time, there is widespread agreement that the borders ignored existing ethnic, cultural, and linguistic divisions, creating tensions that persist today (Rodney 1981; Pakenham 1991; Abraham 2007; Alesina et al. 2011; Michalopoulos and Papaioannu 2016; Christensen and Laitin 2019). In fact even a casual observation of the borders reveals that they often follow longitudinal and latitudinal straight lines.

Regarding South-East Asia, there was no equivalent of the Tordesillas Treaty and of the Berlin conference, that is there was no formal throughout agreed division among European powers, so the settlement patterns were simply driven by colonial expansion, and economic interests, often through commercial companies such as the British East India Company and the Dutch East India Company. Once again, the balance of power among European states,

together with the maritime power and technology, determined the assignment of these territories and, based on the available historical evidence, and there is no reason to believe that the identity of the local population played a key role in this division.

Summarizing, colonization was shaped by the evolution of European military superpowers away from the Tordesillas equilibrium, driven by the relative decline of Spain and Portugal and by the the rise of powerful, Protestant, commercially aggressive, challenger states, and not by the characteristics of the colonized population. There is actually the possibility that the new winners in this European struggle for power were actually able to seize the best territories, although this assumes a deeper knowledge of the colonies characteristics that we cannot assume to be true in general. But even in this case, i.e. even if that systematic knowledge existed, this will not challenge our identification that features colonizer fixed effects.

7 Results

We now discuss the core empirical results of the paper (section 7.1) on the relationship between colonizer-colony folklore similarity and modern income per capita, and their robustness (section 7.2).

7.1 Folklore Similarity and Income per Capita

The main empirical results for the country-level database are summarized in Table 2. First, we find a positive association between median folklore similarity and the modern levels of GDP per capita in the colony, with an elasticity of the effect slightly below 0.2. In terms of the economic significance of the result, moving from the first quartile of deep folklore similarity between colonizer and colonized, equal to 0.037 (Spain and Cuba, for example), to the third, equal to 0.194 (Portugal and Brazil, for example), predicts about a 100% difference in modern GDP per capita at PPP. Thus the effect is economically quite significant.

Second, there is a negative and statistically significant association between the standard deviation of folklore similarity and modern GDP per capita, with an elasticity around 2%. According to the estimates, moving from the median level of folklore standard deviation, equal to 0.025 (Cote d'Ivoire) to the third quartile, equal to about 0.1 (Burkina Faso), results in a

60% lower level of modern GDP per capita (note that the first quartile of folklore standard deviation is equal to 0, reflecting former colonies with only one indigenous ethnic group).

Very similar results obtain from the ethnic level sample, where we find a positive relationship between folklore similarity and cell level income, with an elasticity of the effect of folklore similarity on income between 0.08 and 0.15 depending on the model specification (remember that there is no standard deviation in this model because there is only one ethnic group in the cell over which income in computed). The results are summarized in Table 3.

Controlling for genetic (not available for the disaggregated dataset[12]), linguistic and religious distance between the ex-colony and the ex-colonizer delivers the same empirical results in both empirical specifications. This last result suggests that the effect of cultural similarities that we are estimating is specific to the cultural traits in the oral traditions, as determined by ancient cultural diffusion and exchanges, and not by similarities in other cultural traits (see Berezkin 2015 and D'Amato and Russo 2025-I for a discussion).

Overall, the empirical results that ex-colonies and territories with more similar oral traditions to the colonizers, and with a lower dispersion of oral tradition similarity, are characterized by higher levels of income per capita in the present times. In the remainder of this section, we will extensively discuss the robustness of these empirical results (section 7.2), and also report some additional empirical evidence. The uninterested reader may skip directly to section 9, where we discuss the relationship between folklore similarity and the "Reversal of Fortune" hypothesis.

7.2 Folklore Similarity and Income per Capita: Robustness

In this section we report the results from several robustness checks, especially for the country level sample (country fixed effects, in the cell-level database, already control for many unobservable, time-invariant, potential confounders with country-level variation). To simplify the presentation, we only discuss the results without reporting tables, which are available upon request.

Settlers Colonies. Colonies can be classified in two broad categories, settler colonies,

[12]Genetic distance data are available only for a small set of ethnic groups in the former colonies and, in the present context, the sample size, with this additional control, is too small to estimate the model.

characterized by a large migration from the colonizer, and extraction colonies, with a more limited migration. This distinction is important because, as argued in Acemoglu et al. (2001), the type of institutions that were functional to settler colonies, namely those that encouraged entrepreneurship, are more functional to long-term growth than those there were functional to extraction colonies, that instead fostered rent-extraction. Migration, in turn, was mostly influenced by the geo-climatic conditions of the colony. To control for systematic differences in these two types of colonies, we included, in the country-level regression specification (country fixed effects in the ethnic-level dataset already control for this effect), a dummy in case of settlers colonies, finding robust results for both the Λ and S variables.

The reason why the settler dummy is not part of the baseline set of controls is that it is potentially a "bad control": the decision to migrate and settle, could have been influenced by the diffusion of information on the local population decades after the establishment of the colonial relationship. The coefficient on the settlers dummy is positive and statistically significant, consistently with Acemoglu et al. (2011). We also tried splitting the sample according to the settlers colony dummy, finding similar empirical results as the benchmark in both subsamples, so the effect that we highlight in our analysis is not specific to one of the two broad colony types. The coefficient on the median folklore similarity, however, is much bigger in the settlers colonies sample, while the estimated coefficient on the variability measure has about the same magnitude in both subsamples.

Adding settlers mortality (available for a smaller subsample of ex-colonies) from Acemoglu et al. (2001) instead of the settlers dummy also delivers similar empirical results to the benchmark, and the coefficient on mortality itself is negative and statistically significant, in line again with Acemoglu et al. (2001).

Modern Fractionalization and Polarization. One potential concern with the baseline regression is that the variability measure S might simply pick-up the effect of modern ethnic fractionalization on income (Alesina et al. 2003), which is the case if historical ethnic fragmentation in the colony is correlated to modern ethnic fragmentation. To check for robustness, we included two further controls in the regression: the ethnic fractionalization index computed using population share by ethnicity in 2000, and the number of distinct ethnic groups present in the country as of 2000. We find robust results, with, in particular, a regression coefficient

on S that is very close to the baseline estimate in table 2. We also find robust results to the benchmark when adding ethnic polarization (Estebban and Ray 1994), instead of ethnic fractionalization, using the same population shares by ethnicity in 2000 for the computation. This last result is important because polarization is related to the unrest and civil conflict (Montalvo and Reynal-Querol 2005; Arbatli et al. 2020), an important mechanism for the interpretation of the results, that will be further discussed in later sections.

Cosmological Motifs. So far, we used folklore similarity measures computed using only motifs in the A group of the Berezkin folklore catalog, i.e. those related to "Sun and Moon". The rationale is that they are considered as the oldest motifs in the oral tradition and, hence, those most likely to be exogenous to subsequent colonization episodes. We considered an extension to motifs in the B group, related to the "Origins and characteristics of the environment" which, together with the A group motifs, they make up a full set of cosmological motives. So the resulting folklore similarity and variability measures computed on this extended set can be interpreted as measures of cosmological similarity and variability across populations. The empirical results turned out to be the same as the benchmark, both in the country-level analysis and in the ethnic group level analysis. This is important because, as we will see later on, it helps in the interpretation of the empirical results and of the mechanisms behind them, especially those related to cultural changes induced by colonization, such as the conversions to Christianity (see section 9).

Other Pre-existing Conditions. We argued, following the literature, that cultural similarity between the colonies and the colonizer and its dispersion are an important pre-existing condition in a colonial relationship with an effect on long-term outcomes. To isolate their effect, however, it is important to control for other pre-existing conditions, on the top of those already considered. In the baseline country-level regression, along with local geo-climatic controls, we used the log of the country-level population density in 1500 (McEvedy and Jones 1978) as a generic control. In the cell-level dataset, we considered instead country fixed effects along geo-climatic factors. To further control for the presence of pre-existing conditions in the colonies, we considered three additional controls. The first is the indicator of the political organizational complexity of the indigenous groups from the Ethnographic Atlas, namely the number of jurisdiction levels beyond the local community (see Michalopoulos and

Papaioannu 2013, among others). The second is the indicator of the settlement complexity of the indigenous groups, also from the Ethnographic Atlas. Both variables are available at the ethnic group level so, in the country-level regressions, we simply compute a median over all ethnic groups with centroid within the modern borders of the colony. The third control is the State Antiquity indicator of Bockstette et al. (2002) in 1500 AD (although the results are robust when using antiquity at 1000AD, 0AD, 1000BC and 2000BC), which measures the extent to which complex states where already present at some historical dates within the modern borders of the former colonies, which is available at the country level and, therefore, used as an additional control in the country-level dataset only. In all these cases, we find robust results with respect to the benchmark, that is a positive effect of folklore similarity on long-run income per capita and a negative effect of the variability of folklore similarity. Along the same lines, we found robust results when controlling for historical, country-level, technology adoption in 1500 in the country-level sample, using data from Comin et al. (2010).

Migratory Distance. We included, as an additional control in the regressions, either the migratory distance between the ex-colony capital and Addis-Ababa, in the country-level sample, or the migratory distance between the ethnic group centroid and Addis-Ababa, in the ethnic group level sample. This is important because population with a common migratory history are also characterized by similar oral traditions (Galor et al. 2023; D'Amato and Russo 2025 I), and the migratory distance predicts within-population genetic variability (the "Out-of-Africa" hypothesis), that has been shown to be significantly related to long-run growth (Ashraf and Galor 2013). We found robust empirical results, both for the Λ and S measures. This robustness exercise excludes the possibility of spurious empirical results through the genetic variability/within-group diversity channel, especially, but not limited to, the sample at the country level.

As already stressed in section 3.2, common ancestry, for two ethnic groups, is among the determinants of folklore similarity (See also D'Amato and Russo 2025 I), and common ancestry has been showed to be related to differences in income per capita, mostly because of an effect through technological transfer (Spolaore and Wacziarg 2009). In the baseline regression, we already control for weighted average genetic distance to account for this potential confunding factor. Here we consider an additional control, looking at differences in migratory distance

from Addis-Abeba between the ethnic groups in the colonies and the colonizers. We simply compute medians of the areas of the migratory triangles introduced in D'Amato and Russo (2025 I), which are defined, for an ethnic group pair of which one is in the colony and another in the colonizer, by their migratory distance from Addiss-Abeba and by their distance. Smaller areas are for ethnic groups pairs with a common migratory history and, hence, with a more recent separation from the closest comment ancestor. When adding this further control in the country-level regression, we found robust results.

Legal Origins. We estimated both the country-level and the ethnic-level empirical models with legal origins dummies for the colonizer instead of colonizer fixed effects. This robustness check is important because it allows us to isolate the role of cultural similarity from that of the legal origins. We distinguish the British empire of common law legal origins, from the other colonizers with civil law traditions and we simply code a dummy for common law. We still find a positive and statistically significant coefficient on folklore similarity Λ and a negative and statistically significant coefficient on S, with magnitude in line with the benchmark results. The coefficient on the Common-law legal origins is itself positive and statistically significant in both samples, which aligns with previous evidence on institutional transfer as an effect of colonialism (La Porta et al. 2008).

Duration of the Colony. Modern colonization started in different periods for different areas, with starting dates ranging from the beginning of 1500 (Brazil and Perù, among others) to the end of the 1800 and beginning of 1900 (Kenya and Morocco, among others). Colonization regimes, on the other hand, span from few decades (Morocco and Uganda, among others) to several centuries (India and Nigeria, among others). The reason why it is important to account for these differences is that the role of pre-existing cultural similarities might have been different in these different waves and across regimes of different duration. Think, for example, about the role of religion in modern European states in the XVI century and in the XIX century.

We constructed two variables to account for potential differences in colonization legacies stemming from the different historical periods of colonization. First, we included the country-level total number of years of colonization as additional control, since the length of the colonization period can be related to economic and political long-run outcomes (Grier

1999; Feyrer and Sacerdote 2009; Olsson 2009), especially in case of extractive institutions that drain material and human resources from the colony (La Porta et al. 1999; Sokoloff and Engerman 2000; Acemoglu et al. 2001). We found robust results for both the Λ and the S measures of cultural similarity. Moreover, in the country-level sample, the coefficient of the duration of the colonial period on both modern GDP per capita and GDP per capita is positive and statistically significant.

We also found robust results when including, instead of the total duration, the start and end date of the colonization period. We found a negative and statistically significant coefficient on the starting date, meaning that earlier colonized countries are territories are the onse with a higher GDP per capita today, which is also likely to be related to the difference between the original colonization of the Americas with the historically more recent scramble for Africa.

As a further test, we constructed a colonial dummy equal to one in case, at a pre-specified historical date, the country was under colonial rule. Then we the product of this dummy with the number of years of colonization experienced until that date. The resulting variable can the interpreted as a "Colonial Antiquity" indicator. We choose 4 dates, 1800, 1850, 1900, and 1950, and we use the resulting variables as additional controls in the regression. The result is that the estimated coefficients on both the Λ and the S indicators of folklore similarity remain in line with the benchmark models.

In conclusion, the evidence is that longer colonization periods, that started earlier, are associated with higher modern levels of income per capita in the former colonies, but the role of folklore similarity emerges from the data even when controlling for the duration of the colonial rule.

A Broader Perspective: Asian Empires. The analysis, so far, has been focused on European colonization of overseas territories, mostly through naval conquest. To check if our result can be generalized to other historical episodes of colonization, we expanded the database to four ancient Asian empires that exerted significant influence over broad territories: the Ottoman, the Russian, the Safavid, and the Quing[13]. Whenever we have countries that

[13]The Ottoman empire, established around 1299, spread from North Anatolia to much of the Middle East, Caucasus, Europe and North Africa at its peak. The Safavid empire, founded around 1500 in modern day Iran, expanded its domain from Mesopotamia to Caucasus, central Asia and to the Persian Gulf. The relatively more modern Russian empire, established around mid 1700, at its maximum expansion covered a vast territory from eastern Asia to Europe and Caucasus. The Quing empire of the Manchu, spread, in the second half of

were part of more than one empire, such as Georgia that, at different times, was part of the Ottoman, Russian and Safavid, we compute averages over similarities with all external conquerors, since our goal is to estimate an average effect of a foreign domination on long-run outcomes, which entails assuming that the outcomes are the result of all past dominations without taking a stand on which one is more important.

The main difference between the European empires and Asian empires is that the territorial expansion of the latter was mainly to contiguous areas, based on the power of their land armies, rather than naval power. Given the geographic proximity with the conqueror, and given the importance of human migrations in much of the areas where the empires expanded (something that can be reasonable excluded for the case of the colonies in the benchmark sample, at least in relatively modern times), there is the possibility of endogeneity of the country ethnic composition to unobservables such as the historical development stage, with potential effects on income per capita today. So the results from this robustness test must be taken with caution.

When adding the additional set of countries or ethnic groups within the domain of the four Asian empires to the main samples, we find the same result as in our baseline regressions: median folklore similarity with the external power/conqueror predicts higher levels of income per capita today, while more variability of folklore similarity predicts lower levels of income per capita today, even when conditioning on several covariates including conquerors/colonizers dummies.

Further Robustness Checks. We find robust results when including, as additional controls in the country-level sample, the ex-colony land suitability to agriculture, the ex-colony average terrain ruggedness and the average elevation. The rationale for including these additional bio geographic variables variables is that they all capture finer geo-climatic conditions that can potentially influence the administration of the colonial rule. It is especially important in our case, because it further controls for a pre-existing condition, as highlighted by an extensive literature on how early transitions to agriculture, eased by favorable geo-climatic conditions, helped jump-starting a process of growth and institutional development

the 18th century, from China to large part of eastern Asia, including southeast Asia, and part of central Asia as well.

with long-lasting effects (Ashraf and Galor 2011; Putterman 2008; Olsson and Hibbs 2005, among others). Note that including the country-level neolithic transition timing itself, instead of its bio-geographic determinants, also delivers similar results. Part of the reason why this additional controls do not alter the benchmark results is because population density in 1500 was considered in all regressions, which may already capture the early effects of neolithic transition on development.

In the baseline country-level sample, given the rather small number of observations, we consider robust standard errors without clustering. To test for robustness, we considered, as alternatives, standard errors clustered at the level of the colonizer which turned out to be smaller than the robust standard errors that we report in the tables. Similarly, also when clustering at the level of the continent in which the colony is located results in smaller standard errors. Still in the country-level sample, we also find robust results when using, to measure geographic distance, the median geodesic distance between the centroids of the indigenous groups in the ex-colony from those in the colonizer, instead of the geodesic distance between the country capitals. In fact the difference between these two distances is non-negligible only for the ex-colonies with a significant extension and with scattered ethnic groups, such as the United States, Canada, Brazil, Argentina, the Republic of Congo and India.

The results are also robust if we consider a different criteria for the assignment of former colonies to the colonizers: by assigning Canada to the British empire, rather than the French empire, and by assigning Rwanda to Belgium instead of Germany, also, by excluding Ireland and Malta from the sample, which are the only colonies in Europe, and thus rather different from the rest of the sample, also delivers robust empirical results.

8 Additional Results on other modern outcomes

Civil Conflict and Political Stability. We test now for the possibility that a less conflictual colonial experience determined by more cultural similarity with the colonizer led to politically more stable states post-independence.

We use information from two sources. The first is the UCDP/PRIO armed conflict database, from which we gather information on civil conflicts that took place within the

former colonies between 1946 and 2023. Starting from the raw data, we compute the total number of distinct conflicts in the period, according to the PRIO classification[14]. In both cases, we exclude, from the computation, conflicts with involvement of a foreign state, either supporting the opposition groups or openly joining the conflict.

The second source of information is the World Governance Indicators (WGI) database from the World Bank, which aggregates survey information provided by a large number of different subjects, including country experts, companies, and NGOs, on the perceived quality of the institutions and of governance in the country. We take, in particular, the aggregate scores attributed to "Political Stability", averaged over the entire period for which the score is computed, from 1999 to 2004, and standardized.

In all regressions we include, together with the baseline set of controls used throughout the analysis, a dummy equal to one in case of oil-producing colonies. Moreover, we consider a set of additional control that previous works showed to be related to the probability of civil conflict: ethnic fractionalization (Alesina et al. 2003; Fearon and Laitin 2003) and polarization (Montalvo and Reynal-Querol 2005; Esteban et al. 2012), and within-population diversity proxied by the migratory distance from Addis-Ababa (Arbatli et al. 2020). The regression results are shown in Table 4. Overall, there is evidence that former colonies with an indigenous population more culturally similar to the former colonizers are characterized by fewer post-independence civil conflicts and by more post-independence political stability. Moreover, more variability of folklore similarity with the former colonizer is associated with more post-independence civil conflicts and by less political stability.

In section 10.2, we will report evidence that higher folklore similarity with the colonizer, and lower variability of it, are associated to a higher probability of a direct colonial rule. Direct colonial rules, in turn, are more likely to create stronger, unitary administrative structures, thereby also fostering identification with the central government, whereas indirect rules are associated with weaker structures (Lange 2004 and 2009; Mamdani 1996; Iyer 2010; Michalopolous and Papaioannu 2013; Bolt and Gardner 2020; Blanton et al. 2001). The

[14]For robustness, we also computed the total number of years in the period 1946-2023 in which at least one civil conflict took place. This indicator is different simply because there are both civil conflicts that span more than one year, and years characterized by more than one conflict, that is governments fighting with different opposition groups in separate occasions. The empirical results, however, turned out to be the same as the one discussed in the paper for the number of civil conflicts.

results described in this section are therefore in line with these former works.

Life Expectancy and Institutional Quality. We checked whether folklore similarity between the colonizer and the colony affects other long-run socio-economic outcomes over and above income per capita. For data availability, we conducted the analysis in the sample at the country level only.

First we studied the effect on life expectancy, using data from the World Bank Development Indicators (averaged between 1990 and 2005). Using the same benchmark empirical specification, we find a positive and significant coefficient on median folklore similarity (Λ) and a negative and statistically significant coefficient on the variability measure S. Note, however, that better life expectancy might simply be a consequence of higher income per capita.

Next, we considered several additional indicators of institutional quality from the World Governance Indicators (see above): regulatory quality, government effectiveness, rule of law and control of corruption. Furthermore, we use data on the protection against expropriation from the Political Risk Services, in line with the analysis by AJR (2001). We found that larger folklore similarity between the ex-colonizer and the ex-colony is associated with better regulatory quality, better government efficiency, better rule of law, better control of corruption, and more protection against expropriation. The results are summarized in table 5. As for the variability of folklore similarity, we did not find statistically significant coefficients except for political stability, for which we have a negative sign. We did not find any relationship with the indicator of voice and accountability.

9 Folklore Similarity and the Reversal of Fortunes

Before moving and to pave the way to the exploration of the reasons behind the results provided in the previous sections we investigate here the role of cultural similarities for the reversal of fortunes, which is evidence that colonization affects historical development of former colonies. Acemoglu et al. (2002) documented how the ex-colonies that were relatively richer at the time of colonizations are now poorer, and trace this effect to the impact on institutions. Mahoney (2010) argues for the central importance of relative positions of nations within the overall hierarchy of development in the study of colonialism. Following up on their contribu-

tions, and based on reasons similar to those in support of our main hypothesis, we argue that the reversal is less likely in case of higher cultural proximity of the colonized populations to the colonizer.

While still using the population density in 1500 as a proxy for pre-colonial development (McEvedy and Jones 1978), as in Acemoglu et al. (2002), we develop a new test for the reversal hypothesis, that also overcomes the difficulty of having two, not directly comparable, measures for the historical and modern development stage. First, we rank the ex-colonies from richer to poorer in a global comparison with all other countries, according to population density in 1500, then we rank them according to GDP per capita in 1995, and then compute the difference in ranking. A negative ranking difference measures a reversal, that is a lower position in the modern ranking as compared to the pre-colonial one (for instance, a country moving from the 3rd place in the ranking in 1500 to the 10th in 1995 has a rank difference value of -7). We then code a dummy equal to one in case of reversal, that is of a downward shift in the ranking and zero otherwise. We then regress it on folklore similarity, using the benchmark set of controls, including colonizer and continent fixed effects, and the percentage of the country in tropical and sub-tropical zone, to control for geographic factors that reduced the relative advantage of the tropical territories (the so-called temperate-drift hypothesis). We also include the Neolithic transition timing as a control because of the evidence in Olsson and Paik (2020), who find a reversal for the regions that first completed the transition to agriculture.

We find a negative and statistically significant coefficient on folklore similarity (Λ), meaning that the reversal of fortunes effect is less likely to be observed for ex-colonies with higher folklore similarity with the colonizer. We also find a positive and statistically significant coefficient on the S measure, so reversal is more likely for countries with more volatility of folklore similarity within empires. The results are summarized in table 6. Note that the result is the same in case of a different coding of the dummy, in particular- for example- in case it is equal to one in case of a loss of at least 5 or 10 positions (note that the more we increase this threshold, the more difficult it is to estimate the model given the small number of 1 in the dependent variable).

The result is robust if we add the settler colonies dummy, which is important because the

reversal effect should be more pronounced for exploitation colonies that were mainly used to extract resources. In fact, as shown in column 3 of Table 6 the coefficient on the settlers dummy in the probit regression is negative and statistically significant, consistently, among others, with Acemoglu et al. (2002).

In conclusion, the data suggest that the probability of a reversal of fortune due to colonization, in the terminology by Acemoglu et al. (2002), is lower in case of higher similarity in folkloric traditions between the colonizer and the colonized and higher in case of more variability of the similarity.

10 Mechanisms

We explore now three mechanisms behind the relationship between the colonizer-colonized cultural similarity and long-term development. We start from the impact of cultural similarity on the extent of assimilation of the colonizer's institutions (Section 10.1), as it emerges from the former colonies constitutions at independence. Next, we study the effect of cultural similarity on the administration of the colony, focusing, in particular, on the type of colonial rule enforced (Section 10.2). Finally, we investigate the impact of cultural similarity on conversions to Christianity and, since christian missions typically involved schools, also on education at independence (section 10.3).

Importantly, we do not find any effect of cultural similarity in the colonial relationship on immigration from the colonizer to the colony. In particular, regressing the modern share of the population in the colony made up of ethnic groups with centroids within the border of the colonizer on median folklore similarity, variability, and baseline controls, we do not find statistically significant coefficients on folklore similarity and variability. Thus, the evidence we provide highlights different channels compared to both Putterman and Weil (2010) who documented the effect of post-colonization migration on income per capita, and Acemoglu, Johnson, and Robinson (2011), who focus on settler mortality and migration to explain long-term development of the former colonies.

10.1 Institutional Assimilation

The basic idea that we explore in this section is that, in the colonies, building and sustaining similar institutions to the metropole was easier in case of similar cultural traits of the local population, and more difficult in case of more variability of cultural similarity, over and above other factors that might have shaped institutional design such as the suitability of the environment to settlement by Europeans[15].

We use information from the Comparative Constitution Program (Elkins and Ginsburg 2025), which provides a very rich database of both modern and historical characteristics of the constitutions for most countries. These characteristics include, among others: the presence of an elected head of state, head of government and parliament, together with the electoral rules, quorums and dismissal procedures, if any; the presence of earning disclosure for elected officials, the procedures used to fund political campaigns, and the limits to money used for them; the legislation approval procedures, the supreme court powers and appointment procedures; the organization of courts and the status of judges; the central bank goals; the procedures to declare war and to approve treaties; the duties of both the citizens and the states, if explicitly mentioned; citizenship rights, and eventual explicit protections or discrimination of certain groups; expropriation and eminent domain provisions; civil rights such as free speech, self-determination, and strike; de jure gender equality in matrimonial relationships; the right to bear arms; environmental protection.

We consider the characteristics of the constitutions at independence of the former colony, or as close as possible to that date, with a maximum of 5 years after it. The aim is to explore the influence of colonialism on formal institutions in the sharpest possible way, abstracting from other potential confounders shaping the post-independence constitutional development. We use the entire set of indicators from the database, and build a synthetic measure of constitution similarity between the former colony and the colonizer. It measures, for each pair of a colonial relationship, the relative number of equal constitutional characteristics, as follows:

[15] Acemoglu et al. 2001. On the relationship between culture and institutions see the framework laid out in Acemoglu and Johnson 2025, which provided the inspiration for the exercise we describe in this section. On the relationship between culture and legal origin see also the discussion in La Porta et al. 2008, p.311.

$$C_{KL} = \frac{1}{I^{KL}} \sum_{i=1}^{I^{KL}} \mathbb{1}_{\{y_i^K = y_i^L\}} \tag{6}$$

where y_i^K is the constitutional characteristic i for country K, colonized by country L, at independence, I^{KL} is the total number of characteristics for which there is information for both countries, and $\mathbb{1}$ is the indicator function. So C_{KL} is a measure of the *de jure* constitutional similarity between a colony K and a colonizer L, measured at independence.

We then regress this indicator of constitutional similarity at independence on folklore similarity and controls. The results, summarized in Table 7, highlight that a higher median folklore similarity Λ is associated with more similar constitutions at independence, and that more variability of cultural similarity S is associated with less similar constitutions at independence. The results hold when controlling for the identity of the colonizer, for the pre-colonial development stage and, importantly, for settlers colonies and for pre-colonial state antiquity. As expected, the coefficient on the settler colony dummy, in these regressions, is positive and significant, stressing that institutional transfer was easier for those colonies (as in Acemoglu et al. 2001). We also tried including the colonial rule dummy in the regression, even if it should clearly be considered as a bad control, finding robust empirical results and a non-significant coefficient on the colonial rule variable itself.

Overall, we find robust empirical evidence of an easier transfer of the formal traits of political institutions of the colonizer in colonies where cultural similarity was higher and the group of colonies more homogeneous. This we consider indeed strong evidence consistent with the general theoretical framework outlined in Acemoglu and Johnson (2025).

10.2 Colonial Rule

After the exploration of the characteristics of the formal (de jure) institutions at independence, we explore now the characteristics of the political institutions during the colonial period. Colonialism is, by its very essence, a form of external political domination by a foreign entity with a reduction of, or a strong influence on, sovereignty by local political entities. However the degree of this political domination, that is the extent to which those foreign powers de facto interfered, controlled, or suppressed local authorities, changed markedly both within

and across colonizers. At one extreme, there is the so-called direct colonial rule (Cooper 2005; Mamdani 2018), characterized by an extensive central political and administrative control by the colonizer on the colonial society, with administrators and officers having their origin or descent in the metropole, and little or no role of the local population and institutions. Indirect colonial rules, on the other hand, were characterized by a much weaker control by the colonizer, often through local governance structures, leaders and administrators of local origin that maintained their positions and authority, and local institutions formerly unchanged.

Consistently with our baseline regressions, our goal here is to investigate the within-colonizer variation of the colonial rule. The simple idea is that larger cultural similarity decreases the enforcement cost of a direct rule, since it entails less misunderstanding in the negotiations, similar preferences, a more similar system of beliefs and enforcement by social sanctions, overall a lower probability of a conflict or the resort to violence as a dispute resolution mechanism (D'Amato and Russo 2025 II). Therefore, we expect that more similar cultural traits between the population of the former colony and the colonizer should lead to a higher probability of observing a direct rule. Moreover, larger diversity among the native groups may increase the cost of adopting a direct rule for the simple reason that a single governance structure is more difficult adapted to diverse circumstances, but also larger diversity may increase the benefits from direct rule according to the "divide et impera" principle which makes a direct control easier to enforce: more culturally divided subjects are weaker and make the establishment of a direct rule more convenient for the colonizer.

To get information on the type of colonial rule, we start from the indicator of the "Form of Political Domination" from the Colonial Transformation Dataset[16], which records several aspect of the social, economical, and political changes induced by historical colonization. In the original database the index ranges from 1 in case of "semi-colonialism", which is a form of reduction of local sovereignty by means of non-reciprocal, unequal treaties (forced trade openings or trade interdiction, residency restrictions, change of diplomatic traditions etc.) often imposed as a consequence of war, to 2 in case of an "indirect rule with little influence in internal affairs", 3 in case of an "indirect rule with strong influence in internal affairs", and 4 in case of a "direct rule". Since there are no countries classified in group 1 in our sample of

[16]Data and documentation freely available at: https://www.worlddevelopment.uzh.ch/en/research.html

colonies (the few countries with a 1 in the original sample are not part of the colonial empires that we cover, such as Saudi Arabia, Japan and Thailand), and since the distinction between the 2 and 3 categories is not well grounded among the experts, as acknowledged by the data coder themselves, we decided to code a dummy equal to one in case of a direct colonial rule in the colonies and zero otherwise. Importantly, since the the original database does not cover all the former colonies in our sample (those in the Americas), we expanded it using the same discriminant criteria of the original database and various information sources, including Lange et al. (2006) and information from The Oxford History of the British Empire.

We use a Probit regression specification with both colonizer and continent fixed effects. Among the controls, we add the index of state antiquity, which is important because the presence of pre-colonial institutions affects incentives and constraints to transfer those of the colonizer ones (although we do not find evidence for this channel). We find that median folklore similarity Λ is associated with a higher probability of a direct colonial rule, while more variability of folklore similarity S is associated with a lower probability of a direct rule. The results are summarized in Table 8, and they hold also when controlling for pre-colonial state antiquity and for the settlers colony dummy.

In terms of the effect of the colonial rule, an indirect rule, by fostering local identities, exacerbates ethnic tensions, segregation and conflicts, hindering the formation of a national identity (Ali et al. 2019) and the provision of public goods (Alesina et al. 1999). Furthermore, with a direct colonial rule, the institutional transfer is easier (see section 10.1 for further details) thereby filling the gap between the colonizer and the colony. Note that regressing modern income per capita on the colonial rule dummy, using the same set of baseline controls in the benchmark regressions, including colonizer fixed effects, delivers a positive and statistically significant coefficient on the dummy. So we conclude that there is indeed evidence that former colonies administered through a direct rule tend to have better long-run outcomes.

10.3 Religion and Education

As argued before, the colonial rule entailed changes in the societal fabric, beliefs and behaviors by populations in the former colonies. We now dig deeper into whether precolonial cultural similarities in the colonial relationship affected some important characteristics of the society

in the former colony. We focus on religion, exploring conversions to Christianity, and on education.

Conversions to Christianity. One possible mechanism through which common cultural traits between the colonizer and populations in the colony might have made the colonial relationship less coercive is through an induced change in the system of beliefs in the colonized society. In particular, our hypothesis is that, where missionaries found a larger pool of common ancestral beliefs with the local indigenous population, conversions were easier. Hence we should observe a larger prevalence in the population of those individuals and groups changing their system of religious beliefs, by adopting those of the colonizer, in colonies with more similar cultural traits. Similarly, we should observe less conversions in case of more variability of beliefs.

To test this hypothesis, we use the data on the conversions to Christianity from the Colonial Transformation Dataset, assembled from various sources, such as the World Christian Handbook and the World Christian Encyclopedia, both of which exploit the periodical reports of activities by missionaries. The final database consists of a country-level score ranging from 0, in case of no conversions, to 4 in case of a majority of the indigenous population converted. We run an ordered Probit regression of this conversions score on the folklore similarity measures with the same set of controls as in the benchmark regressions, including colonizer fixed effects that, among other features, controls for the differences in Christian denominations. We find that larger median folklore similarity between colonizer and colonies Λ is associated with a higher conversion score, while more variability of folklore similarity S to lower conversion score. The results are reported in column 2 of Table 9.

Moralizing Gods. To further explore the relationship between folklore similarities and Religion, we use data on the characteristics of gods in the traditional religion of the colonized society from the Ethnographic Atlas. In particular, we look at the variable[17] "High Gods", which records the presence of moralizing gods in the ethnic group mythology according to a score that can take 4 possible values: 1 in case of no gods present, 2 in case of otiose gods, 3

[17]Note that, since information about High Gods in the Ethnographic Atlas is available only for few of the ethnic groups in the colonies, we cannot compute meaningful country level averages to test whether the presence of High Gods might have facilitated conversions, although we believe that this is one of the possible channels.

in case of active gods who do not support morality, and 4 in case of active gods supporting morality of human actions. According to Nunn (2018) information from the Ethnographic Atlas refers to the characteristics of the groups prior to European contact, so we take this data as information about their belief prior to any encounter with Christians from the Old World. We then built a sample of ethnic group pairs and regressed the absolute value of the difference between this high gods score between two groups and their similarity in the deep folklore motifs, controlling for the geodesic distance between their centroids, for linguistic and religious distance, and for a dummy in case of centroids within the borders of the same modern-day country. We found a negative and statistically significant coefficient: larger folklore similarity is related to more similar mythologies as recorded in the ethnographic atlas. The results are summarized in Table 10. Since all colonizers in our sample are Christians, so with a moralizing god, this result can be interpreted as a correlation between deep folklore similarity and the presence of moralizing gods in the traditional mythology of the colonized. Thus this result provides evidence that the effect of folklore similarity on conversions is due to the presence of moralizing gods among the belief system of the colonized.

This last result also opens to the possibility of an additional mechanism linking folklore similarity to long-term outcomes. In particular, religion has often been used to legitimize power ("Divine legitimization"), and moralizing gods, that intervene in human affairs punishing non-compliance (Bentzen and Gokmen 2023) - without punishments, there is no social advantage in institutionalizing religion - ease the divine legitimization of power. Therefore it is plausible that the administration of the colony was easier to legitimize, thereby resulting in less confrontational stances, more integration, less resort to violence and better long-run outcomes, in case of the presence of high gods in the traditional set of beliefs of the indigenous population.

Education. We now study the effect of colonization on education (Bolt and Bezemer 2009; Wantchekon et al. 2015). Our hypothesis is that the accumulation of human capital through formal schooling in the colonies was easier in case of larger cultural similarity with the colonizer. This is related to the conversion channel discussed above, because, by facilitating the spread of Christianity, common cultural traits might have fostered the participation to schools in missions, thereby spreading literacy (perhaps just to better read and understand

the "Word of God") and education in general. In fact Okoye and Pongou (2024) find higher education level, within Nigeria, in the areas with more missionary activities. Similarly, Valencia Caicedo (2019) finds higher educational attainments in areas with Jesuit presence within South America, and Waldinger (2017) reports similar results for Mexico. We also build[18] on Becker and Woessman (2009), who show that bible-reading fostered literacy and the accumulation of human capital in protestant economies, leading to long-run growth, and on Grier (1999), who provides empirical evidence on the effect of colonization on education in Africa. Mokyr (1990) also argues that extractive colonial institutions, and limited investment in education, determined lower technological and institutional progress, hampering long-term growth.

For this exploration we use historical data on the average years of education from CLIO-INFRA, assembled by van Leeuwen, and van Leeuwen-Li (2015). We select, from the database, the year closest to independence of each former colony, in order to have information about the effect of colonization on educational attainment. We regressed this value on median folklore similarity with the colonizer, variability of folklore similarity, and controls, using the same baseline econometric specification with both colonizers and continent fixed effects. As shown in column 4 of Table 9, we find a positive and significant coefficient on folklore similarity, although the coefficient on folklore variability is not statistically significant.

Overall, there is evidence of higher educational attainment in the colony with similar beliefs to the colonizers that promoted the accumulation of literacy and numeracy in the former colonies already at independence, improving their subsequent accumulation of human capital and their performance.

11 Conclusion

We proposed the use of a new, folklore-based, index to measure pre-existing cultural similarities between the colonizer and the colonized in order to study the long-run effects of colonization. We found that the randomly assigned cultural similarity between the people of

[18] Cagé and Rueda (2016) also show that the diffusion of the printing press through protestant missions in Africa is associated with higher education levels. Similarly, Calvi et al. (2022) show higher human capital levels in areas with more protestant missions.

the colonizing country and those in the colonized territories, measured using motifs in the oral traditions of these groups, are associated with a higher level of income per capita today, and with a lower probability of a "Reversal of Fortunes" among the former colonies. We extensively argued in favor of an identification assumption based on the quasi-random assignment of cultural similarity to colonizers using historical information. Our argument relies, in particular, on the absence of detailed ethnographic information at the time of the division of the American, African, and Asian territories among Europeans driven by national interests and based on the balance of power among them.

We also showed colonization represented a transformative historical event for many aspects of the social life of the individuals living in those territories. In particular, in places where the local population was culturally closer in pre-existing traits recorded starting from their oral traditions, everything else equal and conditionally on the identity of the colonizer, there is a higher probability of having a direct colonial rule, and a lower overall violence, thereby easing transfer of institutional traits as measured by the similarity of the political constitutions at independence. Moreover, we showed that cultural similarity affects social and institutional transformation of the populations in the former colonies by affecting the system of beliefs, educational investment and human capital accumulation: higher cultural similarity predicts more conversions to Christianity, better education, and a faster demographic transition.

References

Abraham, Garth. 2007. " "Lines upon Maps": Africa and the Sanctity of African Boundaries", *African Journal of International and Comparative Law* 15:61–84.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. "The Colonial Origins of Comparative Develpment", *American Economic Review* 91(5): 1369-1401.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution", *Quarterly Journal of Economics* 117(4): 1231-1294.

Acemoglu, Daron, and Simon Johnson. 2005. "Unbundling Institutions", *Journal of Political Economy* 113(5): 949-999.

Acemoglu, Daron, Johnson, Simon, Robinson, James. 2005. "The rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth". *American Economic Review* 95(3): 546–579.

Acemoglu, Daron. 2025. "Culture, Institutions and Social Equilibria: A Framework", *Journal of Economic Literature* 63(2): 637-692.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg. 2003. "Fractionalization". *Journal of Economic Growth* 8: 155-194.

Alesina, Alberto, William Easterly, and Janina Matuszeski. 2011. "Artificial States", *Journal of the European Economic Association* 9(2): 246–277.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions", *Quarterly Journal of Economics* 114(4): 1243–84.

Ali, Merima, Odd-Helge Fjeldstad, Boqian Jiang, and Abdulaziz B. Shifa. 2019. "Colonial Legacy, State-building and the Salience of Ethnicity in Sub-Saharan Africa", *The Economic Journal* 129(619): 1048–1081.

Ang, James B. 2013. "Institutions and the Long-run Impact of Early Development", *Journal of Development Economics* 105: Pages 1-18,

Arbatlı, Cemal Eren, Quamrul H. Ashraf, Oded Galor, and Marc Klemp. 2020. "Diversity and Conflict". *Econometrica* 88(2): 727-797.

Asad, Talal. 1973. 'Antropology and the Colonial Encounter', Ithaca Press, Ithaca, NJ.

Ashraf, Quamrul H., and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch", *American Economic Review* 101(5): 2003–2041.

Ashraf, Quamrul and Oded Galor. 2013. " The 'Out of Africa' Hypothesis, Human Genetic Diversity and Comparative Economic Development". *American Economic Review* 103(1): 1-46.

Becker, Sascha O. and Ludger Woessmann. 2009. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History", *The Quarterly Journal of Economics* 124(2): 531–596.

Bentzen, Jeanet Sinding and Gunes Gokmen. 2023. "The power of Religion", *Journal of Economic Growth* 28: 45–78.

Berezkin, Yuri. 2015. "Folklore and Mythology Catalogue: Its Lay-out and Potential for Research". *The Retrospective Methods Network Newsletter* 10: 58–70.

Berezkin, Yuri. 2023. "Comparative Mythology and the Human Past ". *Le Gai Sçavoir (Archaeopress) : 305–314* .

Berque, Jacques. 1962. "Le Maghreb Entre Deux Guerres.", Editions du Seuil, Paris.

Bertocchi, Graziella e Fabio Canova. 2002. "Did Colonization Matter for Growth?: An Empirical Exploration into the Historical causes of Africa's Underdevelopment" *European Economic Review* 46: 1851-1871.

Blanton, Robert, T. David Mason, and Brian Athow. 2001. "Colonial Style and Post-Colonial Ethnic Conflict in Africa", *Journal of Peace Research* 38(4): 473-491.

Blouin, Arthur, and Julian Dyer. 2021. "How Cultures Converge: An Empirical Investigation of Trade and Linguistic Exchange", University of Toronto Working Papers no.691.

Bockstette, Valerie, Areendam Chanda and Louis Putterman. 2002. "States and Markets: the Advantage of an Early Start" *Journal of Economic Growth* 7: 347-369.

Bolt, Jutta, and Dirk Bezemer. 2009. "Understanding Long-Run African Growth: Colonial Institutions or Colonial Education?", *Journal of Development Studies* 45: 24–54.

Bolt, Jutta and Leigh Gardner. 2020. "How Africans Shaped British Colonial Institutions: Evidence from Local Taxation", *The Journal of Economic History* 80(4): 1189-1223.

Bortolini, Eugenio, Luca Pagani, Enrico R. Crema, Stefania Sarno, Chiara Barbieri, Alessio Boattini, Marco Sazzini, Sara Graca da Silva, Gessica Martini, Mait Metspalu, Davide Pettener, Donata Luiselli, and Jamshid J. Tehrani. 2017. "Inferring Patterns of Folktale Diffusion Using Genomic Data". *Proceedings of the National Academy of Sciencies* 114(34): 9140-9145.

British Colonial Office. 1925. "Education Policy in British Tropical Africa", London, UK.

Bruhn, Miriam and Francisco A. Gallego. 2012. "Good, Bad, and Ugly Colonial Activities: Do They Matter for Economic Development?" *The Review of Economics and Statistics* 94 (2): 433–461

Cagé, Julia, and Valeria Rueda. 2016. "The Long-term Effects of the Printing Press in Sub-Saharan Africa", *American Economic Journal: Applied Economics* 8(3): 69-99.

Caicedo, Felipe V.. 2019. "The Mission: Economic Persistence, Human Capital Transmission and Culture in South America." *The Quarterly Journal of Economics* 134(1): 507-556.

Calvi, Rossella, Lauren Hoehn-Velasco, and Federico G. Mantovanelli. 2022. "The Protestant legacy: Missions, Gender, and Human Capital in India", *Journal of Human Resources* 57(6): 1946-1980.

Chanda, Areeendam, C. Justin Cook, and Louis Putterman. 2014. "Persistence of Fortune: Accounting for Population Movements, there was no Post-Columbian Reversal", *American Economic Journal: Macroeconomics* 6(3): 1-28.

Christensen, Darin and David D Laitin. 2019. "African States since Independence: Order, Development, and Democracy", Yale University Press, New Haven, CT.

Comin, Diego and Bart Hobijn. 2004. "Cross-Country Technological Adoption: Making the Theories Face the Facts". *Journal of Monetary Economics* 51(1): 39-83.

Comin, Diego, William Easterly and Erik Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics* 2: 65-97.

Cooper, Frederick. 2005. "Colonialism in Question - Theory, Knowledge, History", University of California Press, Berkley, CA.

D'Amato, Marcello and Francesco Flaviano Russo. 2025 (I). "Cultural Doorways in the Barriers to Development", *Journal of Economic Growth*, forthcoming.

D'Amato, Marcello and Francesco Flaviano Russo. 2025 (II). "Brave Old World: On the Historical Roots of Interstate Conflict", *CSEF Working Paper* no. 749.

De Acosta, José. 1590. "Historia Natural y Moral de Las Indias".

Dundes, Alan. 1980. "Interpreting folklore". Indiana University Press, Bloomington, IN.

Easterly, William. 2006. "The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good", Penguin Books, New York, NJ.

Easterly, William and Ross Levine. 2016. "The European Origins of Economic Development" *Journal of Economic Growth* 21: 225-257.

Elkins, Zachary and Tom Ginsburg. 2025. "Characteristics of National Constitutions, Version 5.0", *Comparative Constitutions Project*.

Engerman, Stanley L., and Kenneth L. Sokoloff. 2002. "Factor Endowments, Inequality, and Paths of Development Among New World Economies", NBER working paper no. 9259, National Bureau of Economic Research, Cambridge, MA.

Ertan, Arhan, Martin Fiszbein, and Louis Putterman. 2016. "Who Was Colonized and When? A cross-country Analysis of Determinants" *European Economic Review* 83: 165-184.

Esteban, Joan, Laura Mayoral, and Debraj Ray. 2012, "Ethnicity and Conflict: An Empirical Study", *American Economic Review* 102: 1310–1342.

Fabian, Johannes. 1983. "Time and the Other: How Anthropology Makes its Object", Columbia University Press, New Yor, NJ.

Fearon, James D. and David D. Laitin 2003. "Ethnicity, Insurgency, and Civil War", *American Political Science Review* 97: 75–90.

Fenske, James, Bishnupriya Gupta and Anwesh Mukhopadhyay. 2025. "Colonial Persistence", CEPR Discussion Paper No. 20135. CEPR Press, Paris & London. https://cepr.org/publications/dp2013520135.

Feyrer, James and Bruce Sacerdote. 2009. "Colonialism and Modern Income: Islands as Natural Experiments", *Review of Economics and Statistics* 91(2): 245-262.

Ferguson, Niall. 2003. "Empire: How Britain Made the Modern World", Penguin, London, UK.

Galor, Oded, Marc Klemp and Daniel C. Wainstock.. 2023. "The Impact of the Prehistoric Out of Africa Migration on Cultural Diversity", NBER working paper no. 31274, NBER, Cambridge, MA.

Gerring, John, Daniel Ziblatt, Johan Van Gorp, and Julian Arevalo. 2011. "An Institutional Theory of Direct and Indirect Rule." *World Politics* 63(3): 377–433

Grier, Robin M.. 1999. "Colonial Legacies and Economic Growth". *Public Choice* 98: 317–335.

Hariri, Jacob Gerner. 2012. "The Autocratic Legacy of Early Statehood", *American Political Science Review* 106(3): 471-494.

Hailey, Baron. 1938. "An African Survey. A Study of Problems arising in Africa South of the Sahara", Oxford University Press, Oxford, UK.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

Iyer, Lakshmi. 2010. "Direct versus Indirect Colonial Rule in India: Long-Term Consequences", *The Review of Economics and Statistics* 92 (4): 693–713

Landes, David S.. 1998. "The Wealth and Poverty of Nations: Why Some are so Rich and Some so Poor". Norton & Co, New York, NJ.

Lange, Matthew. 2004. "British Colonial Legacies and Political Development", *World Development* 32(6): 905-922.

Lange, Matthew. 2009. "Lineages of Despotism and Development", Chicago University Press, Chicago, IL.

Lange, Matthew, James Mahoney and Matthias vom Hau. 2006. "Colonialism and Development: A Comparative Analysis of British and Spanish Colonies", *American Journal of Sociology* 111(5): 1412-1462

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1999. "The Quality of Government", *Journal of Law, Economics, and Organization* 15(1): 222–279.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins", *Journal of Economic Literature* 46 (2): 285–332.

Lee, Alexander, and Kenneth Schultz. 2012. "Comparing British and French Colonial Legacies: A Discontinuity Analysis of Cameroon". *Quarterly Journal of Political Science* 7(4): 365–410.

Letsa, Natalie Wenzell and Martha Wilfahrt. 2020. "The Mechanisms of Direct and Indirect Rule: Colonialism and Economic Development in Africa", *Quarterly Journal of Political Science* 15(4): 539-577.

Lugard, Frederick J.. 1922. "The Dual Mandate in British Tropical Africa", William Blackwood and Sons, London, UK.

Lyautey, Louis Hubert Gonzalve. 1927. "Paroles d'action, Madagascar–Sud-Oranais–Oran–Maroc (1900-1926)". A. Colin, Paris.

Mahoney James . 2010. "Colonialism and Postcolonial Development: Spanish America in Comparative Perspective", Cambridge University Press, Cambridge, UK.

Mamdani, Mahmood. 2018. "Citizen and Subject: Contemporary Africa and the Legacy of Late Colonialism", Princeton University Press, Princeton, NJ.

Matsumura, Katsuo. 2019. "Theories of Diffusionism: Myth and/or Science?" *Comparative Mythology* 5(1): 44-54.

McEvedy, Colin and Richard Jones. 1978. "Atlas of World Population History", Facts on File, New York, NJ.

Michalopoulos, Stelios, and Elias Papaioannou. 2013. "Pre-Colonial Ethnic Institutions and Contemporary African Development" *Econometrica* 81(1): 113-152.

Michalopoulos, Stelios, and Elias Papaioannou. 2016. "The Long-Run Effects of the Scramble for Africa." *American Economic Review* 106 (7): 1802–48.

Michalopoulos, Stelios and Melanie Meng Xue. 2021. "Folklore", *Quarterly Journal of Economics* 136(4): 1993-2046.

Mokyr, Joel. 1990. "The Lever of Riches: Technological Creativity and Economic Progress", Oxford University Press, Oxford, UK.

Montalvo, José G., and Marta Reynal-Querol. 2005. "Ethnic Polarization, Potential Conflict, and Civil Wars", *American Economic Review* 95(3): 796-816.

Nemet, Gregory, Jenna Greene, Finn Mueller-Hansen and Jean C. Minx. 2023. "Dataset on the Adoption of Historical Technologies Informs the Scale-up of Emerging Carbon Dioxide Removal Measures". *Communications Earth and Environment* 4, 397.

Nordhaus, William D.. 2006. "Geography and Macroeconomics: New Data and New Findings", *Proceedings of the National Academy of Sciences* 103(10): 3510-3517.

North, Douglass C.. 1988. "Institutions, Economic Growth and Freedom: An Historical Introduction," in Freedom, Democracy and Economic Welfare. Michael A. Walker, ed., Vancouver: Fraser Institute.

Nunn, Nathan. 2010. "Religious Conversions in Colonial Africa", *American Economic Review: Papers & Proceedings* 100: 147-152.

Nunn, Nathan. 2018. "Ancestral Characteristics of Modern Populations" *Economic History of Developing Regions*, 33 (1): 1-17.

Olsson, Ola. 2009. "On the Democratic Legacy of Colonialism", *Journal of Comparative Economics* 37(4): 534-551.

Olsson, Ola, and Douglas A. Hibbs Jr. 2005. "Biogeography and Long-Run Economic Development", *European Economic Review* 49(4): 909–938.

Olsson Ola, and Christopher Paik. 2020. "A Western Reversal Since the Neolithic? The Long-Run Impact of Early Agriculture", *The Journal of Economic History*, 80(1): 100-135.

Pakenham, Thomas. 1991. "The Scramble for Africa", Avon Books, London, UK.

Padró I Miquel, Gerard, and Pierre Yared. 2012. "The Political Economy of Indirect Control", *The Quarterly Journal of Economics* 127(2): 947–1015.

Pascali, Luigi. 2017. "The Wind of Change: Maritime Technology, Trade, and Economic Development", *American Economic Review* 107(9): 2821-2854.

Pemberton, Trevor J., Michael DeGiorgio, and Noah A. Rosenberg. 2013. "Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation". *G3-Genes/Genomes/Genetics*, 3: 903-919.

Perham, Margery. 1962. "The Colonial Reckoning", Knopf, New York.

Pomerantz, Kenneth. 2000. "The Great Divergence: China, Europe, and the Making of the Modern World Economy", Princeton University Press, Princeton, NJ.

Pratt, Marie L.. 1992. "Imperial Eyes - Travel Writing and Transculturation", Routledge, New York, NJ.

Putterman, Louis. (2008). "Agriculture, Diffusion and Development: Ripple Effects of the Neolithic Revolution", Economica, 75(300): 729–748.

Putterman, Louis, and David M. Weil. 2010. "Post-1500 Population Flows and the Long-Run Determinants of Growth and Inequality", *The Quarterly Journal of Economics* 125(4): 1627–1682.

Rodney, Walter. 1981. "How Europe Underdeveloped Africa", Harvard University Press, Cambridge MA.

Sachs, Jeffrey. 2001. "Tropical Underdevelopment", NBER Working Paper No. 8119, Cambridge, MA.

Service, E. R. . 1955. "Indian-European Relations in Colonial Latin America." *American Anthropologist* 57(3): 411–425.

Smith, G. Elliot. 1916. "The Origin of the Pre-Columbian Civilization of America" *Science* 1128: 190-195.

Sokoloff, Kenneth L., and Stanley L. Engerman. 2000. "Institutions, Factor Endowments, and Paths of Development in the New World" *Journal of Economic Perspectives* 14(3): 217-232.

Spolaore, Enrico, and Romain Wacziarg. 2009. "The Diffusion of Development", *Quarterly Journal of Economics* 124 (2): 469-529.

Spolaore, Enrico, and Romain Wacziarg. 2018. "Ancestry and Development: New Evidence", *Journal of Applied Econometrics* 33(5): 748-762.

Stocking, George W., Jr. 1987. "Victorian Anthropology". Free Press, New York, NJ.

Tankwa, Brendon and Lucas Vazquez Bassat. 2025. "Comprehensive National-level Technology Adoption Dataset (1720-2022)", Mendeley Data, V1, doi: 10.17632/4b65hzcpys.1

Thuillard, Marc. 2021. "Analysis of the Worldwide Distribution of the 'Man or Animal in the Moon' Motifs", *Folklore: Electronic Journal of Folklore* 84: 127-144.

Tylor, Edward Burnett. 1871. "Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom", Vol. 2., J. Murray ed.

Valencia Caicedo, Felipe. 2019. "The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America", *The Quarterly Journal of Economics* 134(1):507–556.

van Leeuwen, Bas and van Leeuwen-Li, Jieli. 2015. "Average Years of Education", IISH Dataverse.

Waldinger, Maria. 2017. "The Long-run Effects of Missionary Orders in Mexico", *Journal of Development Economics* 127: 355-378.

Wantchekon, Leonard, Marko Klašnja and Natalija Novta. 2015. "Education and Human Capital Externalities: Evidence from Colonial Benin", *The Quarterly Journal of Economics* 130(2): 703–75.

Weber, Max. 1918. "Politics as a Vocation".

Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.). 2022. "The ASJP Database version 20" (https://asjp.clld.org/).

Witzel, E. J. Michael. 2012. "The Origins of the World's Mythologies", Oxford University Press, New York, NJ.

Table 1: **Folklore Similarity with the Colonizer, Summary**

	Mean	Median	STD	1^{st} qrt	3^{rd} qrt	N
All Colonies	0.134	0.143	0.104	0.037	0.194	110
British Colonies	0.129	0.128	0.116	0.012	0.215	39
French Colonies	0.117	0.111	0.099	0.010	0.185	27
Spanish Colonies	0.145	0.169	0.084	0.089	0.192	23
Portuguese Colonies	0.139	0.183	0.069	0.090	0.194	7
Dutch Colonies	0.177	0.159	0.069	0.088	0.233	4
Other Colonies	0.143	0.124	0.136	0	0.239	10
Within Colonizers	0.478	0.468	0.091	0.405	0.534	8

Notes: Summary statistics of median folklore similarity between the colonizer and the population in the colonies, for all former colonies (first rows) and for former colonies of a specific colonizer. Other colonies refers to: German, Italian, and Belgian colonies. The last row refers to median folklore similarity between the colonizers. N is the number of countries.

Table 2: **Folklore Similarity and Colonization: Countries**

	Log GDPPC in 1995			
	(1)	(2)	(3)	
Folk Sim (Λ)	0.1817***		0.1813***	0.2110***
	(0.0718)		(0.0687)	(0.0722)
Folk Std (S)		-0.1795***	-0.1791**	-0.2112***
		(0.0571)	(0.0559)	(0.0605)
Lang Dist				1.4819
				(2.5909)
Relig Dist				0.1207**
				(0.0581)
Gen Dist				-0.0262
				(0.0745)
R^2	0.418	0.427	0.468	0.491
Obs	98	98	98	98

Notes: OLS regression results. Observations are for ex-colonies. Dependent variable is in column. Folk Sim is the median cosine similarity measure Λ (in logs). Folk Std is the variability of folklore similarity S (in logs). Lang Dist, Relig Dist and Gen Dist are the logarithms of, respectively, linguistic, religious and genetic distance between the ex-colonizer and the ex-colony. All regressions include: the distance between the capitals of the ex-colony and of the colonizer, the absolute value of the latitude of the capital of the ex-colony, the logarithm of the population density in 1500 in the ex-colony, the percentage of the surface of the ex-colony in tropical or subtropical zones, ex-colonizer fixed effects and continent fixed effects. Robust standard errors in brackets. *** significant at 1% level. * significant at 10% level.

Table 3: **Folklore Similarity and Colonization: Ethnic Groups**

	Log Average GCP at PPP			
	(1)	(2)	(3)	(4)
Folk Sim	0.1568***	0.1529***	0.0785**	0.0843**
	(0.0372)	(0.0342)	(0.0363)	(0.0414)
Lang Dist		-2.9038		2.8574
		(4.8374)		(5.4682)
Relig Dist		-0.1662**		-0.1448**
		(0.0681)		(0.0688)
R^2	0.243	0.232	0.147	0.148
Obs	660	610	660	610
Continent FE	√	√		
Country FE			√	√

Notes: OLS regression results. Observations are for ethnic group centroids in the former colonies. Dependent variable the logarithm of the average gross cell product in the area where the ethnic group centroid is located. Folk Sim is the logarithm of cosine measure of folklore similarity between the ethnic group and the former colonizer. Lang Dist and Relig Dist are the logarithms of linguistic and religious distance between the ethnic group and the ex-colonizer. All regressions include: the distance between the capital of the ex-colonizer and the ethnic group centroid, the absolute value of the latitude of the group centroid, the average modern temperature around the ethnic group centroid and colonizer fixed effects. Regressions in columns 1 and 2 feature continent fixed effects and standard errors clustered at the level of the colonizer (in brackets). Regressions in columns 3 and 4 feature country fixed effects and standard errors clustered at the country level (in brackets). *** significant at 1% level. * significant at 10% level.

Table 4: **Folklore Similarity, Civil Conflict and Political Stability**

	Civil Conflict, Number		Political Stability	
	(1)	(2)	(3)	(4)
Folk Sim (Λ)	-0.3305**	-0.3170*	0.1897***	0.1831**
	(0.1574)	(0.1794)	(0.0786)	(0.0799)
Folk Std (S)	0.5191***	0.3898*	-0.1316**	-0.0492
	(0.1701)	(0.2243)	(0.0628)	(0.0697)
R^2	0.408	0.406	0.315	0.406
Obs	83	75	83	75
Controls		√		√

Notes: OLS regression results. Observations are for former colonies. Dependent variable in columns 1 and 2 is the logarithm of the total number of civil conflicts from the PRIO database (1946-2023). Dependent variable in columns 3 and 4 is the political stability indicator from the Worldwide Governance Indicators database, averaged over the period 1996-2024. Folk Sim is the logarithm of the median cosine similarity measure of folklore (Λ). Folk Std is the logarithm of the variability of folklore similarity (S). All regressions include: the distance between the capitals of the ex-colony and of the colonizer, the absolute value of the latitude of the capital of the ex-colony, the logarithm of the population density in 1500 in the ex-colony, the percentage of the surface of the ex-colony in tropical or subtropical zones, a dummy in case of oil producing ex-colonies, ex-colonizer fixed effects and continent fixed effects. Additional controls included in columns 2 and 4 are: ethnolinguistic fractionalization, ethnic polarization, and migratory distance from Addis-Ababa to the ex-colony capital. *** significant at 1% level. * significant at 10% level.

Table 5: **Folklore Similarity and Institutional Quality**

	Contemporary Quality of Institutions in Former Colonies				
	Reg Qual	Gov Eff	Rule	Corr Cont	Prot Expr
	(2)	(3)	(4)	(5)	(6)
Folk Sim (Λ)	0.2361***	0.2018***	0.2189***	0.1294*	0.5613***
	(0.0735)	(0.0596)	(0.0710)	(0.0711)	(0.1889)
Folk Std (S)	-0.0438	-0.0258	-0.0484	-0.1038*	-0.1377
	(0.0545)	(0.0484)	(0.0557)	(0.0574)	(0.1448)
R^2	0.480	0.534	0.476	0.453	0.461
Obs	90	90	90	90	70

Notes: OLS regression results. Observations are for ex-colonies. Dependent variable in column. Reg Qual is the regulatory quality indicator from the WGI. Gov Eff is the government effectiveness indicator from the WGI. Rule is the rule of law indicator from the WGI. Corr Cont is the corruption control indicator from the WGI. Prot Expr is the indicator of protection from expropriation from the Political Risk Services. Folk Sim is the median cosine similarity measure Λ. Folk Std is the variability of folklore similarity S. All regressions include: the distance between the capitals of the ex-colony and of the colonizer, the absolute value of the latitude of the capital of the ex-colony, the logarithm of the population density in the ex-colony in 1500, the percentage of the surface of the ex-colony in tropical or subtropical zones, state antiquity at 1500, ex-colonizer fixed effects and continent fixed effects. Robust standard errors in brackets. *** significant at the 1% level. ** significant at the 5% level. * significant at the 10% level.

Table 6: **Folklore Similarity and Reversal of Fortunes**

	Probability of Losing Positions in Income Ranking		
	(1)	(2)	(3)
Folk Sim (Λ)	-0.6151***	-0.5387***	-0.3824**
	(0.2293)	(0.1974)	(0.1702)
Folk Std (S)		0.2901**	0.5547***
		(0.1335)	(0.1497)
Settler			-0.2299***
			(0.0636)
Pseudo R^2	0.419	0.471	0.611
Obs	77	77	77

Notes: Marginal effects from a Probit regression. Observations are for ex-colonies. Dependent variable is a dummy equal to one in case of losing position in the modern ranking compared to the historical ranking (population density in 1500 and GDP per capita in 1995). Folk Sim is the median cosine similarity measure Λ. Folk Std is the variability of folklore similarity S. Settler is a dummy equal to one for settlers colonies. All regressions include: the distance between the capitals of the ex-colony and that of the colonizer, the absolute value of the latitude of the capital of the ex-colony, and the percentage of the surface of the ex-colony in tropical or subtropical zones, the agricultural transition timing, ex-colonizer fixed effects and continent fixed effects. Robust standard errors in brackets. *** significant at 1% level. * significant at 10% level.

Table 7: **Folklore Similarity and Constitutions**

	Constitution Similarity Index			
	(1)	(2)	(3)	(4)
Folk Sim (Λ)	0.1683***	0.1532***	0.1367**	0.1336**
	(0.0638)	(0.0631)	(0.0611)	(0.0641)
Folk Std (S)		-2.6913**	-3.1584***	-3.8979***
		(1.1669)	(1.1688)	(1.1412)
Settler			0.4707**	0.3437*
			(0.2113)	(0.2125)
Antiq				-0.2231
				(0.5099)
R^2	0.603	0.624	0.649	0.665
Obs	94	94	94	85

Notes: OLS regression results. Observations are for former colonies. Dependent variable is the Constitution similarity index between the ex-colony and the ex-colonizer at independence. Folk Sim is the median cosine similarity measure Λ. Folk Std is the variability of folklore similarity S. Antiq is pre-colonial state antiquity measured at 1500. All regressions include: the distance between the capitals of the ex-colonizer and of the ex-colony, the absolute value of the latitude of the ex-colony modern capital, the percentage of the ex-colony surface in tropical and sub-tropical zones, the logarithm of pre-colonial population density, continent fixed effects and colonizer fixed effects. Robust standard errors in brackets. *** significant at 1% level. ** significant at the 5% level.

Table 8: **Folklore Similarity and Colonial Rule**

	Direct Colonial Rule		
	(1)	(2)	(3)
Folk Sim (Λ)	0.1016***	0.0924**	0.0961***
	(0.0407)	(0.0394)	(0.0399)
Folk Std (S)		-0.0829**	-0.1025**
		(0.0425)	(0.0493)
Settler			0.1841
			(0.1283)
Antiq			-0.1211
			(0.4060)
Pseudo R^2	0.263	0.287	0.324
Obs	95	95	87

Notes: Marginal effects from a Probit regressions. Observations are for ex-colonies. Dependent variable is a dummy equal to one in case of a direct colonial rule. Folk Sim is the median cosine similarity measure Λ in logs. Folk Std is the variability of folklore similarity S in logs. Settler is a dummy equal to one in case of settlers colonies. Antiq is the state antiquity measure compute at 1500. All regressions include: the distance between the capitals of the ex-colony and of of the colonizer, the number of ethnic groups with centroid within the modern boundaries of the ex-colony, the absolute value of the latitude of the capital of the ex-colony, the logarithm of the population density in the ex-colony in 1500, the percentage of the surface of the ex-colony in tropical or subtropical zones, continent fixed effects and colonizer fixed effects.. Robust standard errors in brackets. *** significant at 1% level. ** significant at 5% level.

Table 9: **Folklore Similarity, Religion and Education**

	Conversions		Education	
	(1)	(2)	(3)	(4)
Folk Sim (Λ)	0.3826***	0.3395**	0.2202***	0.2128***
	(0.1568)	(0.1641)	(0.0834)	(0.08288)
Folk Std (S)		-0.2964		-0.0756
		(0.1987)		(0.0954)
R^2	0.465	0.485	0.581	0.575
Obs	59	59	81	81

Notes: Observations are for ex-colonies. Dependent variable in column. Conversions is the score assigned to the share of the population converted to Christianity (see text). Education is the average number of years of education at independence. Folk Sim is the median cosine similarity measure Λ (in logs). Folk Std is the variability of folklore similarity S (in logs). All regressions include: the distance between the capitals of the ex-colony and of the colonizer, the absolute value of the latitude of the capital of the ex-colony, the logarithm of the population density in the ex-colony in 1500, the percentage of the surface of the ex-colony in tropical or subtropical zones, ex-colonizer fixed effects and continent fixed effects. Robust standard errors in brackets. Ordered Probit regressions results (Pseudo-R^2 reported) in columns 1 and 2. OLS regression results in columns 3 and 4. *** significant at the 1% level. ** significant at the 5% level. * significant at the 10% level.

Table 10: **Folklore Similarity and High Gods**

	Difference in EA scores for the presence of "High Gods"	
	(1)	(2)
Folk Sim (λ)	-0.3650***	-0.3338**
	(0.0238)	(0.0321)
Lang Dist		0.0089***
		(0.0009)
Relig Dist		0.6182
		(0.0201)
R^2	0.592	0.613
Obs	108345	58105

Notes: OLS regression results. Observations are for ethnic group pairs. Dependent variable is the Absolute value of the difference between the "High Gods" scores in the Ethnographic Atlas. Folk Sim is folklore similarity (λ). Lang dist is linguistic distance. Relig Dist is religious distance. All regressions include the geodesic distance between the ethnic group centroids and a dummy in case of centroids in the same modern-day country. Robust standard errors in brackets. *** significant at 1% level. ** significant at the 5% level.

Appendices

Appendix A Sample of Former Colonies

In the country-level dataset, we have the following set of former colonies, assigned to colonizers on the basis of the longest and/or most important foreign domination. England: New Zealand, Zimbabwe, Cayman Islands, Sierra Leone, Myanmar, Ghana, Egypt, Anguilla, Jamaica, Sudan, Ireland, Fiji, Malta, Singapore, Tanzania, Malawi, Solomon Islands, Kenya, Bangladesh, British Virgin Islands, Guyana, Tonga, Lesotho, Pakistan, Gambia, Australia, Nigeria, Sri Lanka, Kuwait, Uganda, Zambia, Antigua and Barbuda, Dominica, United States, Eswatini, and India. France: Chad, Togo, Niger, Senegal, Gabon, Madagascar, Djibouti, Cambodia, Dominican Republic, Benin, Comoros, Mauritania, Mali, Lebanon, Burkina Faso, Guinea, Vietnam, Algeria, French Guyana, Central African Republic, Liberia, Tunisia, Canada, Laos, Morocco, Cameroon and Cote d'Ivoire. Spain: Belize, Guatemala, Equatorial Guinea, Haiti, Paraguay, Panama, El Salvador, Chile, Mexico, Argentina, Bolivia, Costa Rica, Philippines, Peru, Nicaragua, Puerto Rico, Western Sahara, Uruguay, Venezuela, Cuba, Honduras, Ecuador, and Colombia. Portugal: Sao Tome and Principe, Mozambique, Guinea-Bissau, Malaysia, Angola, Macao, Brazil, and Timor-Leste. Netherlands: Indonesia, South Africa and Suriname. Germany: Papua New Guinea, Samoa, Burundi and Namibia. Belgium: Congo and Rwanda. Italy: Eritrea, Libya and Somalia. Note that, as discussed in section 7.2, the results are robust across major re-classifications (Canada, Rwanda, Republic of Congo).

Appendix B Identification of the Empirical Model

The main identification assumption in our analysis is the random assignment of colonizers to territories, which resulted in a random distribution of folklore similarity, conditional on observables. Classical accounts (quote needed) in global history link the emergence and consolidation of colonialism to the historical vicissitudes that led to the formation of Nation States in Europe, emerging from internal fights for supremacy within the territories that would become their borders and, subsequently, to the evolution of the geopolitical equilibrium among

those states. Moreover, colonization episodes have already been used as a quasi-natural experiment (Acemoglu et al. 2001 and 2002) that randomly determined the matching of different populations.

So the idea that similarities in oral traditions between European and indigenous populations might have been the criterion on which European countries may have selected the territories over which to expand their domain has not been contemplated in the account by most of the historians. However some concerns may remain that similarities of cultural traits may have evolved as a consequence of other characteristics (such as bio climatic conditions, trade routes and precolonial exchanges) not adequately controlled for in our regressions.

In this section we extensively argue in favor of the assumption that the assignment of a partner in the colonial relationship, in our sample, is orthogonal to the degree of folklore similarity. We organize the discussion around four main themes: the absence of ethnographic studies before or at the time of colonization (section B.1), the historical evidence on the evolution of the colonization process (section B.2), the empirical design (section B.3), and several empirical tests that validate our assumption (section B.4).

B.1 Ethnographic Studies

Essentially, we cannot make the assumption that systematic information about the system of beliefs of the indigenous population was available to whoever took the decision of where to establish a colony. In fact ethnographic studies actually Followed colonization. In greater detail, the first ethnographic study is arguably the "Florentine Codex" by the Franciscan missionary Bernardino de Sahagun, developed in the second half of 1500, after the Spanish conquest. The book consisted in an extensive study of the Aztec civilization, with information about several aspects including myths of creation and trickster episodes, gathered through fieldwork in Mexico. Dated in the same period, there is also the "Historia natural y moral de las Indias" by the Jesuite Jose de Acosta, with wider coverage of the indigenous groups of Mexico and Perù, the "Histoire d'un voyage fait en la terre du Brésil", by the protestant missionary Jean de Lery, focused on a single group in Brazil, and the "Jesuit Relations" focused on the Algonquian and on the Iroquoian groups of North America. At the beginning of 1600, Richard Hakluyt published instead several collections of travel accounts and reports

by english merchants and diplomats from several areas of the World, but the coverage in terms of ethnic group was limited (few groups scattered between the North-American east coast and the Caribbean, some groups in India and Malaysia, and few between Senegal and Guinea), and in-depth analysis of the traditional beliefs and cultures is absent.

B.2 Historical Evidence

Following up the brief discussion in the main text, we provide here a more extensive discussion of the historical evidence in favor of the hypothesis of random assignment of a cultural similarity between colonizers and colonized.

The first piece of historical evidence comes from the so-called *Treaty of Tordesillas*. Right after the great geographic discoveries, in 1494, Spain (Castille crown) and Portugal with the intermediation of the Pope, signed an agreement with the intent to settle future disputes over discovered territories in the Americas. The treaty initially established a papal demarcation line of 100 leagues (about 320 miles) west of the Cape Verde Islands, such that all lands discovered, or to be discovered, to the west of this line would belong to Spain, and lands to the east would belong to Portugal. Additionally, the treaty allowed Spanish ships to sail through Portugal's zone (east of the line) to reach their own territories, as long as they did not stop or trade there. Importantly, Papal approval was not required. The treaty was a bilateral agreement, even if both parties agreed to seek the Pope's blessing (which came in 1506), the treaty was binding based on their own authority, which can also be interpreted as a sign of rising royal power over that of the Papacy. Since Portugal found the arrangement unfavorable, as it prevented them from having a viable route to Asia via the southern tip of Africa (which Vasco da Gama would soon prove existed), there was a further negotiation that moved the line to 370 leagues (about 1,185 miles) west of the Cape Verde Islands. As a result, Portugal was granted the easternmost part of South America, while Spain all the rest. In addition, the agreement secured for Portugal the route around Africa, solidifying its claims to territories and trading posts along the African coast (Angola, Mozambique) and in India, Malacca, and Macau. Clearly the demarcation was absolutely random and, importantly for our purposes, the agreement was not at all based on the characteristics of the discovered colonies. Even the motivation behind the greed shift of the demarcation line is on the basis of sailing routes.

The treaty represented an astonishing act of arrogance from a modern perspective. Two European kingdoms, with only a fragmentary knowledge of the world's geography, claimed the right to divide the entire globe—including continents and civilizations unknown to them—between themselves. It ignored the rights and sovereignty of the millions of people already living in these lands. The treaty Fueled the Age of Exploration and Global Rivalry: A "Scramble" for Claiming Land: The treaty created a race for both powers to send explorers to map and claim lands on their respective sides of the line before the other power could. Notice that geographical ignorance led to conflict: the imprecise measurement of a "league" and the difficulty of determining longitude at the time made the exact location of the line ambiguous. This led to centuries of disputes, particularly over the Moluccas (the "Spice Islands") in Asia. This rivalry eventually led to the Treaty of Zaragoza (1529), which established an anti-meridian in the Pacific. Importantly, the Treaty of Tordesillas established a Precedent for Colonial Law and Papal Authority: The treaty was part of the "Doctrine of Discovery," a legal and philosophical framework that gave Christian nations the right to claim lands inhabited by non-Christians. It set a precedent for using international law and diplomacy among European powers to manage their colonial competition, even as they violently subjugated indigenous populations.

Soon after, the newcomers in the race to colonies, England, France and the Netherlands rejected the Tordesillas treaty's (Papal) authority, exploiting the weaknesses of Spain and created new legal and political justifications for their own empires.

England and France ended their own internal conflicts (e.g., the English Civil War, the French Wars of Religion), allowing them to focus outward and challenge external rivals like Spain. England: Unified under a strong Tudor and later Stuart monarchy, it developed a powerful navy and a culture of privateering (state-sanctioned piracy) against Spanish treasure ships, with figures like Sir Francis Drake becoming national heroes. France: Under powerful ministers like Richelieu and later the absolute rule of Louis XIV, France became the dominant land power in Europe. This allowed it to also project power overseas without fear of being overrun at home. The Dutch Republic gained independence from Spain and became a financial and commercial powerhouse, with a sophisticated banking system and joint-stock companies such as the Dutch East India Company, that gave them a massive economic advantage.

Spain was bled dry by constant warfare in Europe (e.g., the Eighty Years' War against the Dutch, the Thirty Years' War, conflicts with France). Military Defeats: Key events demonstrated that the Spanish Armada was not invincible. The defeat of the Spanish Armada (1588) by England was a psychological and strategic turning point. It proved that England could challenge Spanish naval supremacy and defend itself from invasion. The ongoing success of the Dutch Revolt showed that Spanish military power could be successfully resisted. The vast influx of American silver caused inflation and made its economy dependent on precious metals rather than production.

On the legal side the "Freedom of the Seas" (Mare Liberum) doctrine, proposed by the Dutch jurist Hugo Grotius in 1609, argued that the sea was international territory, and that no nation could claim sovereignty over it or the trade routes that passed through it. The "Doctrine of Effective Occupation", followed, which became, de facto, the new rule governing colonization. It stated that claiming land was not enough; a European power had to actually settle, administer, and defend a territory to have a legitimate claim. This perfectly suited the challengers, who established colonies and trading posts in "unoccupied" or poorly defended areas of the Americas. Spain, in particular, had vast claims but its actual control was often limited to key ports and mining regions in the interior. The coasts, especially in North America and the Caribbean, were vulnerable.

The new colonial powers used a combination of different strategies to gain new territories. First, smuggling and piracy were used, in order to Undermine the Spanish commercial monopoly by trading illegally with their colonies or outright seizing their silver shipments (examples: English "Sea Dogs," Dutch and French privateers in the Caribbean). Second, settling in regions that were technically within the Spanish/Portuguese sphere but were not effectively controlled (examples: Jamestown and Plymouth for England, Quebec for France, New Amsterdam for the Netherlands). Capturing strategic islands and territories to serve as bases for trade, piracy, and further expansion (examples: Barbados and Jamaica for England, Saint-Domingue and Martinique for France, Curacao for the Netherlands). Ignoring the Tordesillas line entirely and focusing on the rich trade of the Indian Ocean and Spice Islands, where Portuguese power was stretched thin (examples: the English East India Company and Dutch East India Company (f. 1602) established trading posts and wrested control of the

spice trade from Portugal).

The treaties that followed (London 1604; Westphalia 1648; Madrid 1670; Utrecht 1713) simply ratified the new balance of power that had already been established by force and settlement. In particular, the Treaty of London ended the Anglo-Spanish War. While it didn't formally renounce Spanish claims, it effectively marked England's successful entry into the colonial game without being crushed by Spain. the Treaty of Westphalia ended the Thirty Years' War. It recognized the independence of the Dutch Republic from Spain, formally acknowledging a major Protestant power that had built a global empire in defiance of Iberian claims. with the Treaty of Madrid, Spain formally recognized English sovereignty over its American conquests, including Jamaica. With the Treaty of Utrecht, at the end of the War of Spanish Succession, a defeated Spain, under the new Bourbon dynasty (tied to France), was forced to make major concessions, most significantly Britain gained the exclusive right to supply African slaves to Spanish America, which gave Britain a legal foot in the door of the previously closed Spanish colonial economy.

For what concerns Africa, until the first half of the 19^{th} century, the European involvement in Africa was mostly limited to coastal trading posts dictated by the convenience on the navigation routes and the dominant winds (Pascali 2017). However the fast developing industrial economies determined both a growing demand for raw materials, and a growing need to find new markets, leading European powers to compete for influence and resources in what is generally known as the "Scramble for Africa". To prevent conflicts, 14 major European powers gathered in Berlin in 1884 to negotiate, without representatives from African nations. The main outcome, together with a re-affirmation of the principle of Effective Occupation, was a "Partition of Africa" into spheres of influence, colonies and protectorates, through the construction of formal boundaries. Given the very limited knowledge of the local population by European at that time, there is widespread agreement that the borders ignored existing ethnic, cultural, and linguistic divisions, creating tensions that persist today (Rodney 1981; Pakenham 1991; Abraham 2007; Alesina et al. 2011; Michalopoulos and Papaioannu 2016; Christensen and Laitin 2019). In fact even a casual observation of the borders reveals that they mainly follow longitudinal and latitudinal straight lines. Therefore we can safely consider the assignment of colonizers to ethnic groups in Africa is exogenous with respect to their

ancestral culture.

For South-East Asia, there was no equivalent of the Tordesillas Treaty and of the Berlin conference, that is there was no formal throughout agreed division among European powers. There are however few exceptions, such as the treaty of Nanking, which established Britain in Hong Kong, and the treaty of Saigon, that gained France control in southern Vietnam, but they are generally the results of violent confrontations with the local population. Thus the settlement patterns were simply driven by colonial expansion, often through commercial companies such as the British East India Company and the Dutch East India Company. The goal, before the industrial revolution, was to control the highly profitable spice trade while, afterwards, to establish new markets and a steady supply of raw inputs for the newly established industries. As for the case of other colonies, the balance of power among European states, together with the maritime power and technology, determined the assignment of territories and, based on the available historical evidence, there is no reason to believe that the identity of the local population played a key role in this division, over and above the physical characteristics of the environment they lived in.

Summarizing, the matching between colonized and colonizes seems to be the outcome of the evolution of the balance of power among european countries, with little to no regard of the identity of the local population in the colonies.

B.3 Empirical Design

One crucial feature of the baseline econometric specification is the inclusion of colonizer dummies. This allows us to control, among other factors, for the specific set of institutions and technologies that could potentially be transferred within the empire, and for the power of the colonizer, which could have determined the colony assignment. In other words, since the history of the colonial expansion was partly shaped by the evolution of the balance of power between European States (see section B.2), there is the possibility of assortative matching, that is better colonies assigned to the most powerful colonizer. Controlling for colonizer fixed effects shields from this potential threat to identification.

There is, in principle, the possibility that colonization, because of frequent exchanges between the colonizer and the colonized, might have led to cultural similarities as a result. For

instance, loanwords due to, say, commercial interactions, probability induced lower linguistic differences (Blouin and Dyer 2021). It is to control for this possibility that our baseline measure of cultural similarity is computed using only the deep folklore motifs related to sun and moon, which are older than all others, thereby excluding the possibility of cultural similarities determined as a result of colonization. In fact, according to Berezkin (2015), such motifs contain the deepest elements of the oral tradition, most of which plausibly date back to about 10000 BC, although dating, in such contexts, must be taken with caution, being based on anecdotal, rather than hard, evidence.

While deep folklore motifs are plausibly exogenous to the colonization process, there still is the possibility that the unobservable determinants of folklore similarity might have had an impact on both colonization and on the current levels of income per capita. For instance, folklorists believe that motifs followed the patterns of ancient demic movements (Bortolini at al. 2017), and perhaps also of ancient technological diffusion, for instance the Neolithic transition. We assume that, whatever this ancient potential determinants, their effect on current income levels is captured by their effect on the pre-colonial development stage. As a benchmark measure for this pre-colonial conditions, we use, in all regression specifications, the population level in 1500, but we also used more explicit indicators of technology adoption and an alternative indicator equal to the differences in the neolithic transition timing. Note also that controlling on pre-existing conditions in the colony also helps controlling for potential selection within colonizer: institutions and technology might have been transferred only on some colonies on the basis on their stage of development. Note also that controlling on pre-existing conditions in the colony also helps controlling for potential selection within colonizer: institutions and technology might have been transferred only on some colonies on the basis on their stage of development.

B.4 Empirical Support to Identification

Overall, we did not find any empirical evidence of selection into colonization driven by folklore similarity. First, if there was selection into colonization, there would be both a high value of folklore similarity between colonized and colonizer, and a small variability of folklore similarity within empires, but this is not the case. First, there are a number of colonies, for each colonizer,

with 0 deep folklore similarity: 6 out of 27 (22%) for French colonies, 9 out of 38 (%) for the UK, but only 2 out of 23 (9%) for Spain. Second, the average folklore similarity between all possible World country pairs that we can construct (using the indigenous population with centroid within the borders) is equal to 0.193, while the average folklore similarity between colonizer and colony in the sample turns out to be actually smaller: 0.125 and the UK, 0.146 for Spain and 0.137 for Portugal. Third, the coefficient of variation of the distribution of deep folklore similarity measured over all possible world country pairs is equal to 70%, whereas the coefficient of variation of the distribution of deep folklore similarity between colonized and colonies is equal to 84% for French colonies, 96.8% for British colonies and 61% for Spanish colonies (Portuguese colonies are too few to meaningfully compute summary stats).

The second, more elaborated, test for selection into colonization consists in computing counterfactual deep folklore similarity between each ex-colony and all the European countries who had colonies, and then compare whether the deep folklore similarity between each ex-colony and its historical colonizer is bigger than the folklore similarity with all other potential colonizers. The answer is yes for only 5 out of 105 former colonies: Angola, Brazil, Cuba, Sri-Lanka and Vietnam (In appendix we provide the results from a further test developed along similar lines).

We also regressed, in a sample of all possible world country pairs, a dummy for a colonial relationship on deep folklore similarity, controlling for geo-climatic factors, for other measures of cultural similarity (linguistic, religious and genetic) and for continent fixed effects. We restricted the sample to couples where one country is one of the European colonizers (France, UK, Netherlands, Spain, Portugal, Germany and Italy) and the other a country outside Europe. The result is a non-significant coefficient on folklore similarity.

Lastly, we computed an average difference between the deep folklore similarity between the colony and all the potential colonizers (that is the European countries which had colonies: France, England, Spain, Portugal, Netherlands, Germany, Belgium, and Italy) and the deep folklore similarity between the colony and the actual colonizer. For a country i colonized by country j, this measure is computed as follows:

$$F_{ij} = \frac{1}{S-1} \sum_{s=1}^{S} \left(\Lambda_{is} - \Lambda_{ij} \right) \tag{7}$$

where Λ_{ij} is the deep folklore similarity between the colony and its colonizer, Λ_{ij} are the folklore similarities with all potential colonizers (note that the measure is scaled by $S-1$ because one of terms in the summation is zero - for $s=j$). We then regressed this measure on colonizer dummies controlling for distance and for the number of colonies (stock measure, over all years in the past). The resulting regression coefficients on the dummies turned out to be not statistically significant: a further support that the results provided in the previous sections are not driven by selection into colonization.

Appendix C Cultural Similarities and Colonial Rule.

The importance of the identity of the local population in shaping the implementation of the colonial rule has been a prominent topic of the political debate Long before these issues were formalized in various modern social sciences. For instance, Bartolomé de las Casas argued that indigenous societies in the Americas were governed by intelligible moral and political orders, and that colonial coercion arose from the refusal to recognize this fact. In the "Brevísima relación de la destrucción de las Indias" (1552), he described indigenous populations as endowed with reason and capable of peaceful self-government, suggesting that systematic coercion was not inevitable, rather it reflected a failure to adapt governance to culturally distinct societies. Similarly, although at the other side of the globe, British administrators repeatedly noted that governance in India required engagement with local customs and institutions. In his "Minute on Indian Education" (1835), Thomas Babington Macaulay observed that British rule operated over populations "whose laws, religion, and social habits are altogether different from our own," implicitly acknowledging limits to the applicability of metropolitan institutions. These concerns were articulated most explicitly by Alexis de Tocqueville in his writings on French rule in Algeria. In his "Travail sur l'Algérie" (1841), Tocqueville argued that Algeria could not be governed through the same institutional framework as metropolitan France because the local "population has customs, ideas, and laws so different from ours that it is almost impossible to apply to them the same system of administration." Notice that in Tocqueville's account, exceptional legal regimes and coercive authority were not moral aberrations but rational responses by the colonizer to the perceived incompatibility between the attempt

to establish new institutions in new territories and local social structures. These historical chronicles, of course, do not provide causal evidence; rather, they motivate the conceptual stance of this paper by showing that informed contemporary observers from the metropole understood cultural similarity as a key factor shaping the governance technology of the empire administration.

Frederick Lugard famously argued that the effectiveness of indirect rule rested on the degree to which colonial authority could be embedded in existing indigenous institutions. In his text on the Dual Mandate, Lugard wrote that "indirect rule rests upon the existence of an ordered native society... where political institutions are weak or unfamiliar, every act of administration becomes laborious and uncertain" (Lugard, The Dual Mandate in British Tropical Africa, 1922). Ethnographers such as Maurice Delafosse argued that France had mistakenly attempted to marginalize traditional authorities, he concludes that effective rule required working "through indigenous institutions." (Delafosse, French Colonialism Unmasked, 2002). These insights informed the turn toward association, a doctrine recognizing that cultural and institutional distance required hybrid administrative forms. The Iberian empires articulated closely related principles much earlier. Francisco de Vitoria argued in 1539 that Amerindian societies possessed "ordered republics" with legitimate laws and magistrates, implying that governance required reckoning indigenous political structures (Vitoria, Relectio de Indis, 1539). The twentieth-century Portuguese "indigenato system" explicitly codified differential legal regimes on the grounds that European civil law was unsuitable for those deemed culturally distant (Estatuto do Indigenato, Portuguese Africa, 1930). The Dutch and Belgian empires, though administratively distinct, expressed the same concerns. J. C. van Leur wrote that European authority in Indonesia was "a thin layer spread over a deeply rooted native political order," and that control became "weaker and more burdensome" where indigenous institutions were fragmented (van Leur, Indonesian Trade and Society, 1955). Belgian administrators warned that governance was viable only when "chiefs whose authority is understood by the population" could be relied upon, while invented chiefs obliged constant administrative intervention (Moulaert, Organisation du Congo Belge, 1913).