NBER WORKING PAPER SERIES

FOLKLORE

Stelios Michalopoulos
Melanie Meng Xue

Working Paper 25430
http://www.nber.org/papers/w25430

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2019

We are extremely grateful to Yuri Berezkin for generously sharing his lifetime work on folklore classification and clarifying many of its aspects. We would like to thank Benjamin Enke, Rafael La Porta, Ernesto Dal Bo, Roland Bénabou, Paul Seabright and Max Winkler for their insightful comments. Seminar and workshop participants at George Washington, Bocconi University, Università Cattolica del Sacro Cuore, NBER Political Economy Meeting, New Economic School, New York University, Simon Fraser University, Tel Aviv University, Toulouse IAST, UC Berkeley, Universitat Pompeu Fabra, University of Chicago, University of Connecticut, and University of Zurich offered many useful suggestions. Rohit Chaparala, Masahiro Kubo, Xueyun Luo, Zhihan Wang, and Xiaochen Yan provided superlative research assistance. Melanie Meng Xue acknowledges financial support from Donorstrust, the Balzan Foundation and the Center for Economic History at Northwestern University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Stelios Michalopoulos and Melanie Meng Xue. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Folklore
Stelios Michalopoulos and Melanie Meng Xue
NBER Working Paper No. 25430
January 2019
JEL No. N00,Z1,Z13

ABSTRACT

Folklore is the collection of traditional beliefs, customs, and stories of a community, passed through the generations by word of mouth. This vast expressive body, studied by the corresponding discipline of folklore, has evaded the attention of economists. In this study we do four things that reveal the tremendous potential of this corpus for understanding comparative development and culture. First, we introduce and describe a unique catalogue of folklore that codes the presence of thousands of motifs for roughly 1,000 pre-industrial societies. Second, we use a dictionary-based approach to elicit group-specific measures of various traits related to the natural environment, institutional framework, and mode of subsistence. We establish that these proxies are in accordance with the ethnographic record, and illustrate how to use a group's oral tradition to quantify non-extant characteristics of preindustrial societies. Third, we use folklore to uncover the historical cultural values of a group. Doing so allows us to test various influential conjectures among social scientists including the original affluent society, the culture of honor among pastoralists, the role of family in extended kinship systems and the intensity of trade and rule-following norms in politically centralized group. Finally, we explore how cultural norms inferred via text analysis of oral traditions predict contemporary attitudes and beliefs.

Stelios Michalopoulos
Brown University
Department of Economics
64 Waterman Street
Providence, RI 02912
and CEPR
and also NBER
smichalo@brown.edu

Melanie Meng Xue
Department of Economics
Northwestern University
2211 Campus Drive
Room 3469
Evanston, IL 60208
melanie.xue@northwestern.edu

1 Introduction

Over the last two decades a burgeoning body of work has emerged which explores the deep roots of comparative development.[1] Moving to the subnational level and recognizing the significance of groups (ethnic, linguistic, and religious) has improved our understanding of the process of development. The combination of geographic information systems with the ethnographic record has allowed researchers to test at a large-scale, long-standing conjectures developed by historians, anthropologists, geographers, and evolutionary biologists regarding preference formation, institutional and societal traits, beliefs and attitudes, and their consequences for contemporary economic performance (see Diamond and Robinson (2010), Nunn (2012), Spolaore and Wacziarg (2013), Ashraf and Galor (2017), Michalopoulos and Papaioannou (2017b)).

This renaissance of works exploring the shadow of history on today's outcomes has given rise to a set of critiques. The first one starts with the observation that our understanding of historical cultural values is at best fragmentary. This is especially problematic in the context of works where historical legacies and/or geographical attributes influence present-day economic outcomes via their impact on unobserved historical cultural traits which are in turn assumed to be reflected in contemporary values. Moreover, the absence of proxies of historical attitudes renders inquiries into how norms change and why they persist intractable. Attempts to address these issues have been made in the context of specific traits, regions and time periods, but a comprehensive answer is lacking.[2] A second set of critiques stresses the weaknesses of the ethnographic work of George Peter Murdock (1967), reflected in the *Ethnographic Atlas* (EA), including the incomplete coverage of certain economic and social aspects. This corpus has been fruitfully used by social scientists over the last 20 years. In this study we show how integrating a group's folklore in the analysis can greatly expand the scope of questions asked, open a window into the historical heritage across societies, and improve upon existing approaches.

But what is folklore? Folklore is the collection of traditional beliefs, customs, myths, legends, and stories of a community, passed through the generations by word of mouth. This vast expressive body of culture, studied by the corresponding discipline of folklore, has evaded the attention of economists. Nevertheless, the appreciation that narratives are essential building blocks of our society has gained momentum recently within economics and beyond. At a broad level, Harari (2015) for example, identifies in the common myths and legends that exist in people's collective imagination the defining characteristic of homo sapiens. Within economics

[1] See Michalopoulos and Papaioannou (2017a) for a compilation of many of these seminal studies.

[2] Algan and Cahuc (2010) provide an nice attempt to uncover trust values for most of the 20th century, and a case of persistence of traits has been convincingly shown by Voigtlander and Voth (2012) in the context of anti-Semitism in Germany. Giuliano and Nunn (2017) find that historical climate shocks weaken the contemporary transmission of cultural traits.

Shiller (2017) points out that, despite the central role of stories in motivating and connecting activities to deeply-felt values and needs, our understanding is at best superficial, and calls for quantitative work to shed light on the relationship between narratives and economic fluctuations.[3]

In this study we do four things that reveal the tremendous potential of oral traditions for economists and political scientists interested in comparative development and culture. First, we introduce a unique database of folklore that codes the spatial distribution of thousands of motifs across roughly $1,000$ societies worldwide. This is the lifetime work of the eminent anthropologist and folklorist Yuri Berezkin. The author uses the expressions "folklore," and "mythology" interchangeably to refer to all kinds of traditional stories and tales, long and short, sacred and profane. His catalogue is described in Berezkin (2015a) and Berezkin (2016) where he analyzes the thematic classification and areal distribution of folklore-mythological motifs. But what is a motif? For folklorists a motif is the main analytical unit in a tale/myth. This is any episode or image related to, or described in, narratives in the oral tradition. In section 3 we discuss the relationship between tales and motifs and why folklorists have converged into using motifs in classifying a society's oral tradition. Moreover, we provide details of the structure of Berezkin's corpus.

Second, we link the groups in Berezkin's catalogue to those in the EA, effectively adding the oral traditions to the ethnographic record of these societies. We then employ a dictionary-based approach to elicit group-specific measures of various traits related to the natural environment, the institutional framework, and the mode of subsistence. Among other things, we show that groups whose folklore has a higher intensity of earthquake-related motifs live closer to earthquake-prone regions, those found on fertile lands display more motifs related to agriculture, and finally those residing in areas more often struck by thunderstorms have more motifs rationalizing this natural phenomenon. Besides establishing that salient elements of the physical environment are manifested in a group's oral tradition, we also show that folklore-based measures of political complexity and type of subsistence robustly correspond to the analogous traits in the EA, suggesting the usefulness of folklore in quantifying currently non-extant characteristics of preindustrial societies, including the importance of trade.

Third, we attempt to uncover the societies' historical norms. Specifically, we use two psychosocial dictionaries. The Moral Foundations Dictionary (MFD) as well as the Linguistic Inquiry and Word Count (LIWC) which have been widely used in linguistics, psychology, and sociology. Reconstructing the historical cultural landscape across groups allows us to test various influential conjectures among social scientists including the affluent society among for-

[3] See Bénabou, Falk, and Tirole (2018) for a theoretical investigation of the interplay between narratives and moral reasoning.

agers, the culture of violence and honor among pastoralists, the importance of family members in groups organized along extended family lines, the interplay between political complexity on the one end, and trade and rule-following norms on the other end.

In the last part of the paper we explore how the cultural norms we reconstructed from the groups' oral traditions map into contemporary attitudes and beliefs, reflected in modern-day surveys. In particular, we demonstrate the predictive power of folklore-based measures of culture on current norms regarding tolerance towards violence and anti-social behavior, trust toward family members and attitudes towards enjoying one's time, suggesting that folklore itself may be one of the vehicles via which culture is vertically transmitted across generations.

The rest of the paper is organized as follows. In Section 2 we relate our study to existing works in folklore, culture, historical development, and text analysis and provide a brief history of the field of folkloristics. In Section 3 we introduce the work of Yuri Berezkin. We offer a detailed description of his catalogue, its advantages and potential biases, and how it compares with other existing works in comparative mythology. We also introduce and discuss the dictionaries we use to extract information from folklore. Namely, the ConceptNet, the LIWC and the MFD dictionaries. In Section 4 we detail our empirical approach and present our results in four parts. In Section 5 we conclude offering some thoughts on future work.

2 The Added Value of Folklore

Group-level historical data, contrary to country-level statistics, are scarce particularly for societies outside Europe. The only systematic effort to recover the institutional, economic, and societal makeup for a large cross section of preindustrial societies is the *Ethnographic Atlas (and the Standard Cross-Cultural Sample* which has a wealth of information, albeit for a handful of societies). Synthesizing a large body of anthropological research, George Peter Murdock (1967) and subsequent authors have put together an impressive dataset for a large cross section of ethnic groups around the world. The results of this major effort are recorded in the *Ethnographic Atlas*, reflecting societal, institutional, and economic traits of $1,264$ ethnicities. Thanks to this body of work, the research on the cultural, institutional, and social correlates of growth has moved beyond country boundaries by combining Murdock's EA and the mapping of the spatial distribution of ethnicities (Murdock (1959)) with the underlying location-specific traits to shed light on the origins and consequences of a variety of economic and institutional traits.[4]

Starting with Landes (1998), cultural explanations have been thrust into the spotlight.

[4] See, among others, Nunn and Wantchekon (2011), Michalopoulos (2012), Fenske (2013), Giuliano and Nunn (2013), Osafo-Kwaako and Robinson (2013), Michalopoulos and Papaioannou (2014), Fenske (2014), Alsan (2015), Bentzen, Hariri, and Robinson (2015), Mayshar, Moav, Neeman, and Pascali (2015), Michalopoulos and Papaioannou (2016), Cervellati, Chiovelli, and Esposito (2017), and Michalopoulos, Putterman, and Weil (2018).

For example, a recent addition to the list of candidates on the origins of comparative development is a 2015 book by Harari (2015), who makes the bold claim that the successes and failures of human society are deeply rooted in the common myths that exist in people's collective imagination. Such an intriguing conjecture is hard to assess quantitatively. Along with cultural explanations as drivers of the observed differences in comparative development, a lively debate has emerged regarding the historical interplay between institutions and culture. Nevertheless, progress in these areas has been hindered by the mere fact that value surveys are not available for preindustrial societies.

This paper proposes a way to recover these traits by showing that folklore can shed important light on a group's historical heritage, including beliefs and attitudes of ancestral populations conspicuously absent from the historical record. According to the Oxford English Dictionary, folklore is "The traditional beliefs, customs, and stories of a community, passed through the generations by word of mouth." This very definition of folklore is akin to how economists define culture (Alesina and Giuliano, 2015). Incidentally but importantly for our purposes, folklore is also an academic discipline whose subject matter comprises the sum total of traditionally derived and orally or imitatively transmitted literature, material culture, and customs.

The insights from this discipline have been overlooked within economics. In other social sciences, folklore is gradually being integrated. For example, in a recent paper, Ross, Greenhill, and Atkinson (2013) study the diffusion of a specific folktale and its spatial variation within Europe. The researchers draw insights from population genetics to analyze 700 variants of a folktale ("The Kind and the Unkind Girls") from 31 ethnolinguistic populations with an average of 23 variants each. They find that geographical distance and ethnolinguistic affiliation exert significant independent effects on folktale diversity.

But how do we analyze folklore? In Section 3 we provide a detailed discussion. Broadly speaking, what we have for each society is a set of motifs indicating the presence of a particular theme or an episode in the group's oral tradition. A motif comes with a title and a short (usually two-line) description. For instance, motif $k56$ (present in 172 oral traditions), is titled along the lines of the folktale studied by Ross, Greenhill, and Atkinson (2013), namely: "Kind and unkind girls," and has the following description: "a girl or a woman meets powerful person, behaves herself in a right way and is successful. Another behaves in a wrong way and suffers the opposite (is punished or not rewarded)." This example is deliberately chosen to illustrate the usefulness of a motif as an aggregator of folktale content across multiple variants of a given theme.

This means that text comprises our underlying data. In the social sciences, where the

interest in culture is perhaps most pronounced, the most common type of text analysis examining culture has been, one form or another of, content analysis (Berelson (1952); North (1963); Gebner (1969); Holsti (1969); and Gottschalk and Gleser (1969)). Within this research tradition, the focus has been on concepts and their distribution within and across texts. Over the last few years, text analysis has seen great advances and taken a center stage thanks to the abundance of text (from millions of digitized written sources and online content). For reviews of studies in text analysis in political science and sociology, see Grimmer and Stewart (2013) and Evans and Aceves (2016), respectively. Gentzkow, Kelly, and Taddy (2017) provide an excellent entry into the available text-analysis methods along with their corresponding weaknesses and advantages.

The approach we employ to quantify folklore is the dictionary-based method which connects counts of specific words to latent, unobserved attributes we wish to quantify. This is the simplest and most commonly used method. Besides its simplicity, it is also appropriate for our setting. In dictionary-based methods, one specifies a mapping between the counts of specific words and the latent outcomes. For example, Gentzkow and Shapiro (2010) count the number of newspaper articles containing partisan phrases, whereas Saiz and Simonsohn (2013) enter search queries in Google to obtain document-frequency measures of corruption by country, US states and cities, counting the number of web pages measuring combinations of city names and terms related to corruption.[5]

Our analysis is closely related to the works of Tetlock (2007), Baker, Bloom, and Davis (2016) and Enke (2017), who use a pre-specified dictionary of terms to obtain an estimate of the outcome of interest.[6] In Baker, Bloom, and Davis (2016), the authors use the count of articles in a given newspaper-month containing a set of pre-specified terms such as "policy," "uncertainty," and "Federal Reserve," with the outcome of interest being the degree of "policy uncertainty" in the economy. Similarly, Enke (2017) in order to quantify the extent to which US presidential candidates emphasize universal moral principles relative to "tribalistic" values he applies the bag-of-words from MFD on presidential speeches.

Inspired by these three papers, we follow a similar approach. Specifically, when we want to capture relatively "objective" features of a group, including attributes of the physical environment or the subsistence economy, we use the ConceptNet knowledge graph originated

[5] Gentzkow, Shapiro, and Taddy (2016) apply a structural choice model and methods from machine learning to study trends in the partisanship of congressional speech from 1873 to 2016.

[6] See Antweiler and Frank (2004) for an early text analysis on how messages posted on Internet stock message boards may reflect the views of day traders. Similarly, Bollen and Mao (2011) document a significant link between Twitter messages and the stock market using other dictionary-based tools such as OpinionFinder. Finally, Wisniewski and Lambe (2013) utilize predefined word lists to construct an index of negative media attention toward the banking sector and find that the former Granger-causes bank stock returns during the 2007–2009 financial crisis.

at the MIT Media lab. Specifically, for each concept we are interested in measuring, we use the top-50 related terms that come from the word embeddings built from ConceptNet.

When we want to capture a group's psychological states we rely on dictionaries generated by social psychologists. The LIWC dictionary has been widely used among social scientists. It classifies words into psychologically meaningful categories as well as areas of personal concern. Similarly, the moral foundations theory with its 5 dimensions of morality comes with its corresponding dictionaries. Thanks to their intellectual origins being firmly rooted in the social sciences and the humanities, the resulting bag-of-words classifications by the LIWC and the MFD are likely to be appropriate for the classification of oral traditions. We discuss the pros and cons of the individual dictionaries as well as the dictionary-approach itself below.

2.1 A Short History of Folklore

Folklore studies began in the early 19th century. In 1846, William Thoms invented the word "folklore"to replace existing terms including "popular antiquities." The terms "folk" and "lore" simply refer to "ordinary people" and "knowledge," respectively.[7] The first generation of folklorists focused exclusively on illiterate peasants, and on groups such as the Romani vagabonds, who would be more or less unaffected by the sweeping social changes of the era, and attempted to document their historical customs and beliefs preserved in the oral traditions. The understanding was that folklore reflected the cultural beliefs of ordinary people in opposition to those of the elite. With increasing industrialization, urbanization, and the rise of literacy throughout Europe in the 19th century, folklorists were concerned that the oral tradition, the lore of the rural folk, would be lost. It was thought that the stories, beliefs, and customs were surviving fragments of a cultural mythology of the region, often predating the spread of Christianity.

From an anthropological perspective, folklore is one of the most important components that make up the culture of a given people (Bascom (1953)). Importantly, folklore is considered a key mechanism for preserving a group's tradition. According to Bascom, there are four functions of folklore: informally teaching cultural attitudes, escaping the accepted limitations of a culture, maintaining cultural identity, and validating existing cultural norms.

For over 150 years from the early 19th to the mid-20th centuries, a vast body of work accumulated from collectors in all parts of the world as they listened to story-tellers and, with better and better techniques, recorded and published what they heard. Hence, the very nature of folklore, that is, its transmission via oral storytelling, might appear to be a source of idiosyncratic variability. According to Barre Toelken, however, this was held in check by the forces of orthodoxy and tradition, which were the "twin laws of folklore performance" (Toelken

[7] A more contemporary definition of "folk" is a social group that includes two or more persons with common traits who express their shared identity through distinctive traditions.

(1996)). Audiences expected storytellers to retell familiar stories, and this expectation reined in tendencies to innovate or adapt folklore traditions. To rationalize the stability of the narrative, a famous early 20th-century folklorist, Walter Anderson, posited a double redundancy, that is, a feedback loop between performing and hearing the performance multiple times, in order to retain the essential elements of the tale (Dorst (2016), Fine (1979)).

The natural next step was the development of techniques to categorize this wealth of information for comparative analysis. This was a critical advance in the discipline of folklore, and the indexing of tale types and motifs lies at the heart of its comparative framework. The concept of tale type was first delineated by the Hungarian folklorist Jànos Honti in 1939. Honti proposed three different ways of considering a tale type as a unit of analysis of folklore: first, a tale type consists of a specific set of motifs; second, a tale type does not duplicate with other tale types; third, a tale type manifests itself through multiple existence (termed variants). The motif is defined as "the smallest element in a tale having a power to persist in tradition" (Thompson (1946)). Both are hypothetical archetypes established by comparing a large number of texts that share a common core. The methodology most closely associated with the use of tale types and motifs in comparative mythology is the historic-geographic approach that began in the late 19th century. It focused on establishing an oral tradition, identifying its geographic origin and its spread. The method was questioned and criticized in the wake of postmodernism in the humanities, alongside large paradigm shifts in the discipline. However, it has remained popular as a methodological package for classifying textual sources in comparative analysis, and as a tool for organizing the data according to degree of similarity. In 1982, Richard Dorson declared the historic-geographic method the dominant force in folklore science.

Folklorists working in the historic-geographic tradition often follow the Aarne-Thompson (AT) classification systems. The latter include the AT motif-index, and the AT tale type index which was updated by Uther and is now known as the Aarne-Thompson-Uther classification system (ATU). The ATU classifies $2,239$ tale types and their variants, Uther (2004). The AT motif-index refers to the motif-index of folklore literature created by Stith Thompson in 1955. The AT/ATU system was originally developed to study European oral traditions, limiting its usefulness for classifying folklore from other parts of the world (see the criticism of the classical historic-geographic method by Goldberg (1984)). Besides, ethnic attribution is rarely available for folklore found in the non-Western world, even in several of the major regional indices that followed the ATU index. Although Thompson's motif index partially overcomes the lack of non-European coverage, the distribution of motifs remains skewed towards Europe and obscene-type motifs are intentionally left out, see Dundes (1997).

Yuri Berezkin's Folklore and Mythology Catalogue is a pioneering effort in modifying

and extending the ATU classification system, enabling a global comparative perspective of oral traditions. It is important to keep in mind that there is also the *Encyclopedia of the Folktale (Enzyklopädie des Märchens)* an impressive compilation of almost two centuries of international research in the field of folk narrative tradition. However, it focuses on the oral and literary narrative traditions of Europe and other countries influenced by Europeans. Moreover, there are motif indexes compiled for specific regions that may be more relevant for those who have a regional focus in their research. See El-Shamy (2004), for example, for a classification of the folktales in the Arab world.

3 Berezkin's Folklore and Mythology Catalogue

A critical dimension of the Folklore and Mythology Catalogue is that Berezkin does not limit himself to the European-based ATU tale type classification, or S. Thompson's Motif-index of folklore-literature. To accommodate the richness of non-European folklore, he defines a motif as "any image, structure, element of the plot or any combination of such elements which could be found in at least two (practically, in many) texts including sacred and profane ones." Starting with indigenous societies in the Americas and extending his classification to groups in the Old World over the course of 40 years, Berezkin has compiled a unique database of folklore and mythology for 945 groups worldwide, categorizing their oral traditions into 2,461 motifs. The fruit of his intellectual labor is a unique dataset on oral traditions with an unprecedented global coverage. See Figure 1 for the spatial distribution of groups in Berezkin's catalogue.

Berezkin builds on the historic-geographic method, but with a different goal from its predecessors since he is primarily interested in understanding the historical spread of motifs across societies and is influenced in terms of methods and theory by Boas (1898, 2002) and his students.[8] Classifying non-European and European oral traditions independently, ensures that what is being explored are potential dissimilarities among oral traditions themselves, rather than dissimilarities between European and non-European folklore.[9]

To encode his motifs, he has consulted an impressive list of 6,239 books (67%) and journal articles (33%)[10] from 4,041 authors, published by 4,932 publishing houses in 32 different

[8] An advantage of this approach is that Berezkin retains the historic-geographic method's focus on the formal characteristic of folkloric traditions rather than on subjective attitudes of the narrator and the collector.

[9] The following example illustrates the lopsided coverage of non-European groups in Thompson's motif index. Take two groups the Irish and the Guarani. In the Thompson (Berezkin) index there are approximately 8,000 (268) motifs in the Irish oral tradition compared to 36 (82) recorded for the Guarani.

[10] Here are some folklore journals: The Folk-Lore Journal, Journal of American Folklore, Journal of Latin American Lore, Latin American Indian Literature Journal, The Journal of the Polynesian Society, Journal de la Societe des Americanistes, The Journal of the Anthropological Institute of Great Britain and Ireland, The Journal of Indo-European Studies, Journal de la Societe Finno-Ougrienne, Oral Traditions, etc.

languages.[11] The bulk of the materials in the textual catalogue were published during the 20th century. Appendix Figures 1*a* and 1*b* plot the distribution of the dates of publication across all titles in the catalogue as well as the distribution of the earliest publication consulted per oral tradition, respectively. In Table 1 Panel *A* we report the language of publication across all titles. The most common languages in descending order are English, Russian, Spanish, German, French and Portuguese.

The median group in Berezkin has 59 motifs (see Figure 2 for the distribution of the number of motifs across oral traditions). Russians have the largest number of motifs and only one group has a single recorded motif; the Yeyi in northwestern Botswana. In Table 1, Panel *B*, we report the top 10 groups in terms of motifs. For the average society, Berezkin consulted 15 publications from 12 authors, written in 3.5 languages, and published by 13 publishers.

Broadly speaking, there are two types of motifs, those describing adventures and tricks (roughly 50%) and the remaining are related to cosmology and etiology. Furthermore, Berezkin classifies the 2,461 motifs into 12 categories. These are: (a) Sun and Moon; (b) Moon spots, stars, constellations; (c) Cosmogony, the earth and the sky, etiology of the elements, natural and biological phenomena (fire, water, soil, thunderstorms, dreams, etc.), cataclysms and cosmic threats, spirits of nature; (d) Origin of death, diseases, and hard life; (e) Origin of human beings, ethnic groups, etiology of human anatomy, strange body configuration, ways of behavior, marriages before the establishment of the present norms; (f) Origin and interpretation of culture elements, in particular related to agriculture, inadequate forms of subsistence and economic activity before the establishment of the present norms; (g) Etiology of plants and animals and of their peculiar features, particular animals as protagonists of cosmological stories, metamorphoses, weather, and calendar; (h) Queer and monstrous beings, creatures, objects and loci, folk beliefs related to particular phenomena and objects; (i) Identification of protagonists of the stories with particular animals or persons with particular qualities; (j) Adventures; (k) Tricks and competitions won thanks to deception, absurd, and obscene behavior; and (l) Formulae. In Table 1, Panel *C*, we report one motif in each of these categories.

In the Appendix we describe the construction of one particular motif and the underlying sources of an oral tradition typical of Berezkin's catalogue.

Caveats Naturally, there are several limitations to using this database that one needs to keep in mind. First, one may wonder about the time period the motifs and underlying myths and tales correspond to. Since folklorists were interested in collecting stories that were

[11] Here is an example of a reference on the oral tradition of the Inuit: "Stories of the Black River People. From stories told by Robert (Nasruk) Cleveland of Shungnak (Alaska). Tape-recorded by Don Charles Foote. Transcribed by Minnies (Aliitchak) Gray. Translated and proofread by Ruth (Taqaviñ) Ramoth-Sampson and Angeline (Ipiilik). Newlin Anchorage: University of Alaska (1980)".

relatively untouched by the waves of modernization of the 19th and 20th centuries, the motifs are likely to provide a snapshot of folk life from the preindustrial times. In fact, for the median society in Berezkin's catalogue the earliest publication consulted dates back to 1904. But how far back in history do these motifs go? There is no simple answer to this question. The traditional historic-geographic approach to the tale-type was originally understood as a narrative plot with a more or less precise origin in space and time. However, this idea has been severely criticized by Jason (1970) and Goldberg (1984) and eventually abandoned, recognizing the inherent uncertainty in coming up with convincing estimates.

Nevertheless, it is commonly understood that some motifs likely predate others. For example, cosmological motifs are thought to be significantly older than those regarding adventures and tricks, and Berezkin himself has published a series of studies looking at the areal distribution of individual motifs in relationship to large-scale population movements, migrations, cultural contacts and interactions in history and prehistory (see Berezkin and Duvakin (2016), and Berezkin (2015b) among others). Hence, in the comparisons below with the Ethnographic Atlas, it is useful to keep in mind that the information coded in folklore is potentially mapping into a historical horizon deeper than the date of first publication (around the turn of the century for the median society). Below we will also use some time-varying, location-specific traits to further explore this question.

Second, in his coding of folklore, Berezkin does not code motifs which are universal or found in a single oral tradition. This is also the case for the ATU classification. This is not surprising. The focus of comparative mythology is on motifs that can be found in multiple societies. To the extent that both the observed and the unobserved motifs (i.e., those that are unique to a given society) are drawn from the same distribution of themes and images, our quantification of group-specific traits is defensible.

Third, we do not have a count of how often a given tale or motif is repeated in a society's folklore; that is, we do not know the popularity of a motif nor the number of variants of a particular tale (which can be numerous, as the study by Ross, Greenhill, and Atkinson (2013) reviewed above suggests). Hence, our folklore values reflect the extensive margin of themes in the oral tradition. How much a given oral tradition has been originally studied will naturally influence the number of variants recorded per tale. So, focusing on whether a given motif is present may help mitigate concerns regarding differential sampling across traditions. Similarly, we do not know the number of tales and legends in a given oral tradition since there is no one-to-one mapping between tales and motifs. One tale may map into multiple motifs and vice versa.

Fourth, the definition of a group in Berezkin's catalogue is usually along linguistic lines,

but sometimes he groups related societies when he cannot find enough information on the oral tradition of each individual group.

Fifth, as it will become apparent in the empirical specifications, we often control for country-specific constants. Why are we doing this given the historical nature of most of our exercises? Besides the obvious benefit of accounting for the broad differences in geography and ecology as well as the preindustrial historical legacies across modern-day countries, the inclusion of country fixed effects is primarily motivated by concerns regarding the intensity at which oral traditions were originally recorded by folklorists and subsequently surveyed by Berezkin himself in different parts of the globe. One may worry, for example, that the coverage is systematically different in certain parts of Asia or Melanesia compared to parts of Europe or the Americas. Hence, by including country-specific constants, we hope to mitigate concerns related to unbalanced coverage across countries in terms of the quality and breadth of the underlying recorded oral traditions (assuming that groups within the same country are more likely to be sampled by folklorists during the same time period and presumably with similar biases and available technology).[12] Taking a step further, we will explicitly control for the number of publications in a group's sources.

Sixth, there was substantial heterogeneity in the folklore collection over time. Earlier generations of folklore scholars were often missionaries or hobbyists, whereas later generations were academics including anthropologists who did their field work in small-scale societies. Differences in the timing of being surveyed might reflect differences in the tastes and interests of the original story collectors. To partially account for this, we will control for the year of the earliest publication per oral tradition.

Interpreting Folklore

In the early years of folklore, the main task of folklorists was to collect and classify different folklore materials, paying relatively little attention to its interpretation. Indeed, the primary focus was on the "lore" per se. During the 20th century, understanding the relationship between the "folk" and the "lore" became a focal topic for a long list of preeminent scholars of different backgrounds. While summarizing the various approaches is not our goal, we mention some of them to illustrate Joseph Campbell's prescient point made in 1968 that "there is no final system for the interpretation of myths, and there never will be any such thing."

Broadly speaking, there are three dominant approaches: the humanistic, the anthropo-

[12] A way to gauge the explanatory power of country-specific features on the coverage of oral traditions is to record the R-square from regressing features of Berezkin's underlying references per oral tradition on country fixed effects. Doing so, we find that country-specific constants explain between 30% and 40% of the variation in the number of publications, number of distinct languages in the publication record, and year of first publication across groups.

logical, and the psychological. Myths and folktales have been viewed, among other functions, as reflecting: (i) the customs, institutions and beliefs of a given society, (ii) (distorted) historical facts, (iii) allegories of deeper philosophical issues, (iv) moral truths, (v) explanations of the cosmos, and (vi) metaphors or symbols of the unknown or of religious rituals.

According to the pattern theory of culture, see Benedict (1934), all parts of culture are related and reflect the same values and beliefs, with folklore offering a window into it. Anthropologists have taken an ethnographic approach, a structuralist approach, and more recently, a symbolic anthropology approach to shed light on the function and meaning of folklore, Green (1997).[13] Within psychology, the use of depth psychology, or psychoanalytic approaches, is the dominant method in interpreting folklore. Freud and Jung were pioneers in applying this approach to folklore. Jung approached myths as essentially static symbolism, Jung (1968), where recurring themes or motifs can be attributed to some deeply-seated human needs or conflicts, called psychological archetypes. Hence, the "primitive" mentality is what is "hidden in the unconscious and reflected in folkloric symbols," Green (1997).

Because we work with motifs, i.e., images and episodes of myths, legends, and tales — both of which are intentionally deprived of details — the humanistic approach, which mainly emphasizes the role of the narrator, is less relevant in our context. Between the anthropological and psychological paradigms, we are agnostic. In fact, to maximize the amount of information we may extract from the body of folklore materials, we employ psychosocial dictionaries (the MFD and the LIWC) which have been useful tools for psychoanalysts and cultural anthropologists alike.

Dictionaries

Below we briefly describe the dictionaries we use to classify motifs into different bags of words (BoW).

3.0.1 ConceptNet

ConceptNet is a knowledge representation project, designed to represent the general human knowledge and how it is expressed in natural language. It originated from MIT Media Lab's crowdsourcing project Open Mind Common Sense, and has benefited from other crowdsourced resources and expert-created resources, such as Wikimedia projects, Games with a Purpose, Princeton University's WordNet, DBPedia, OpenCyc, and Umbel. It integrates knowledge from formal and informal sources and makes them available through a common representation

[13] Lévi-Strauss, a proponent of the structural study of myth, sees in myth's structures elementary contradictions or problems which no society can resolve. Hence, in order to understand myths one may focus only on the repeated patterns and their structural relations to each other, Levi-Strauss (1955).

and includes common-sense associations made among concepts by ordinary people. Examples of works that use this knowledge base include recognizing activities of daily living, Ullberg, Coradeschi, and Pecora (2010), and analyzing the emotional content of text, Cambria, Hussain, Havasi, and Eckl (2010).

ConceptNet provides vector representations of the terms that can be compared for similarity. We use the Web API of ConceptNet: api.conceptnet.io. One of the functions available on this API is to use semantic vectors (also known as word embeddings) built from ConceptNet and other inputs to find related terms.[14] From this API, we retrieve the 50 most related English words for each seed word to construct a ConceptNet-based BoW. We use the latter when we want to obtain measures of relatively "objective" features including the physical environment, mode of subsistence and degree of political complexity reflected in the group's oral tradition.

3.0.2 LIWC

LIWC is a text analysis program that counts words in psychologically meaningful categories, see Pennebaker, Booth, and Francis (2007). The development of this computer program first began in the early 1990s. Over time, there have been several versions of LIWC. In this study, we use LIWC2015 which is the latest.

The LIWC2015 Dictionary is composed of almost $6,400$ words, word stems, and emoticons. The construction of LIWC has undergone several steps, from word collection, judge rating phase, base rate analyses, candidate word list generation, psychometric evaluation, refinement phase in addition to summary variables. Several stages of this procedure involve judges to vote on keeping or omitting words, which introduces a human component to quality control.

In this study, we focus on the 41 word categories concerning psychological constructs, together with 6 personal concern categories. The former include affective processes (positive emotion, negative emotion), social processes (family, friends, female, male), cognitive processes (insight, causation, discrepancy, tentative, certainty, differentiation), perceptual processes (see, hear, feel), biological processes (body, health, sexual, ingestion), drives and needs (affiliation, achievement, power, reward, risk), time orientations (past focus, present focus, future focus), relativity (motion, space, time) and personal concerns (work, achievement, leisure, home, money, religion, death).

Tausczik and Pennebaker (2010) summarize hundreds of studies using LIWC demonstrating its ability to detect meaning in a wide variety of experimental settings and the validity of the LIWC categories across a large number of psychological domains.

[14] The embeddings are a version of ConceptNet Numberbatch. In an ACL 2017 SemEval paper, ConceptNet Numberbatch is shown to be the best word embeddings one can use, Speer and Lowry-Duda (2017).

3.0.3 MFD

To get at the moral values of a society's folklore we use the Moral Foundations Theory, developed by Haidt and Craig (2004) and Graham, Haidt, Koleva, Motyl, Iyer, Wojcik, and Ditto (2013) as a positive framework of morality. It was designed to explain the origins, development, and cultural variations of morality. MFT argues that moral intuitions are based on five foundations and linked to adaptive challenges. The five foundations are the following:

(i) Care/harm: This foundation is related to our long evolution as mammals with attachment systems and an ability to feel (and dislike) the pain of others. It underlies virtues of kindness, gentleness, and nurturance.

(i) Fairness/cheating: This foundation is related to the evolutionary process of reciprocal altruism. It generates ideas of justice, rights, and autonomy.

(iii) Loyalty/betrayal: This foundation is related to our long history as tribal creatures able to form shifting coalitions. It underlies virtues of patriotism and self-sacrifice for the group.

(iv) Authority/subversion: This foundation was shaped by our long primate history of hierarchical social interactions. It underlies virtues of leadership and followership, including deference to legitimate authority and respect for traditions.

(v) Sanctity/degradation: This foundation was shaped by the psychology of disgust and contamination.

According to Haidt (2007) these five foundations underlie the cross-cultural differences in moral reasoning. To make cross-cultural comparisons, we use the Moral Foundations Dictionary to construct measures of morality at a society level based on the motifs present in that society. The Dictionary is created by Graham, Haidt, and Nosek (2009) and contains 295 words and word stems.

Discussion The LIWC and MFD are mapping tools. They map a given text to dictionary-supplied categories. Many of these categories may potentially capture meaningful cultural indicators (e.g. "authority/subversion," "family," or "retaliation"). To use these tools, we first break down each motif title and description into words. Then we look up all the words therein and tag the motif to the appropriate category(ies). To arrive at our society-specific estimates, we add up all motifs within each dictionary category to quantify the intensity of a particular concept within a group's oral tradition.

Our dictionary-based approach provides an initial examination of the cultural context of folklore. Computerized text analysis programs, LIWC or MFD, are meant to be probabilistic systems. We may misclassify individual motifs, but as long as the errors are not systematic, we may still obtain a set of cultural proxies at the level of the oral tradition. Despite its

limitations, the dictionary-based method offers several advantages, including the minimization of subjective interpretation of ad-hoc BoW. By focusing on the title and the short description of each motif rather than the tales themselves, we only make use of the essential episode of a story. This largely removes subjective influences introduced by the narrator and the collector.

Nevertheless, the objective nature of this approach also creates its limitations. The LIWC and MFD categories "have proven to supply useful information about a wide variety of texts. But it remains up to the researchers, not the computer, to create knowledge and insight from this mapped information." Such dictionaries are unable to detect context, irony, sarcasm, and idioms. In addition, both LIWC and MFD were developed over the last few decades. Hence, some words are only applicable in the modern context. Given the historical context myths and legends correspond to, it is possible that aspects of oral tradition remain unexplored when viewed through the lenses of these contemporary psychosocial dictionaries.

We recognize that the dictionary-based method is one of the many methods available for text analysis. There are two reasons we employ this one. First, the former is the most commonly used in the social sciences and is simple, transparent to implement, and easy to replicate. Second, Baker et al. (2016), who are interested in measuring the degree of policy uncertainty, use a pre-specified dictionary because there is no ground truth on the actual level of policy uncertainty to develop training data for a supervised model. Also, fitting a model would be unlikely to endogenously detect policy uncertainty as a topic. This reasoning largely applies to our setting. Obtaining reliable training data for complex concepts such as "respect" or "anger" or "leisure", is not straightforward.[15] Having said that we hope that subsequent research by moving beyond dictionary-based methods will enrich our understanding of oral traditions.

4 Empirical Analysis

The empirical analysis proceeds in two steps. First, we assess the predictive power of oral traditions vis a vis an array of observable group-specific traits regarding a society's physical environment and traits known from its ethnographic record. Specifically, for each attribute we wish to quantify, we use the top-50 related words generated by ConceptNet. We show that folklore-based measures of the economy and the society complement our understanding of a group's historical account, improving and extending the set of societal historical traits. In the second step, we employ the psychological dictionaries discussed above to uncover proxies of historical attitudes of the ethnographic societies and use the uncovered norms to shed light on famous conjectures among social scientists. Finally, we explore how these past norms predict

[15] Moreover, topic modeling is unlikely to pick such themes independently.

contemporary beliefs and attitudes.

4.1 Folklore and the Natural Environment

Have groups' natural environment left an imprint on the images and episodes of their oral tradition? Folklorists think that the answer to this question is far from trivial for at least two reasons. First, a society's oral tradition is often thought of as a depository of a group's (migration) history, suggesting that folklore elements can be preserved even if the landscapes, climates, and social configurations in which they are told, have changed (Berezkin (2015a)). Second, it is well understood that folktales and myths travel across space as a result of interactions among individuals belonging to different oral traditions. Eventually through a process known as assimilation, Boas (1916), the borrowed stories are integrated into the stock of the indigenous folklore. These processes of migration and cross-pollination inherent in the formation of a group's oral tradition suggest that its link with a society's landscape at the time of observation (in the late 19th, early 20th century) is not straightforward. With these caveats in mind, we hope that documenting that a group's natural environment is reflected in its folklore images will increase our confidence in using the oral tradition to surmise other aspects of the group for which less is currently known.

We check the following four traits that we think can be reliably measured both in the folklore and in the physical environment of a group. These are: proximity to earthquake zones; intensity of lightning strikes; annual temperature; and caloric suitability for agriculture for the crops available before 1500 CE. See Table 2-Panel *A* for the respective summary statistics across the 945 oral traditions in Berezkin's catalogue. Are these natural phenomena salient enough to manifest themselves in a society's folklore? This is what we explore below.

Specifically, we report OLS estimates of the following type:[16]

$$\ln(1 + \text{Topic-Specific Motifs}_{i,c}) = a_c + \beta \text{Geo}_{i,c} + \gamma \textit{Baseline Controls}_{i,c} + \varepsilon_{i,c} \quad (1)$$

The dependent variable, $\text{Topic-Specific Motifs}_{i,c}$, is the number of motifs mentioning a particular feature of group i, located in country/continent c; a_c reflects continental or country-specific constants; and $GEO_{i,c}$ is a vector of geographic traits. We use the group's centroid (recorded by Berezkin) to compute the distance terms and a radius of 50 kilometers to construct the values of the respective geographic and ecological attributes. The baseline controls are (i) the log number of motifs per group, ln(# Motifs), (ii) the log word-count length of an average motif, (iii) the number of publications that Berezkin has consulted for a group's folklore, and the year of the earliest publication Berezkin has accessed for group i.

[16] Results are similar using Poisson models or using OLS with the level instead of the log transformation of the count of topic-specific motifs.

The rationale for accounting for these features is the following. First, the number of motifs is, in principle, a variable of independent interest, partially reflecting the underlying breadth of themes, images, and episodes present in the oral tradition of a given society. However, the same variable also naturally reflects the intensity with which a given society has been sampled both by folklorists and eventually by Berezkin himself. Controlling for the average number of words per motif is important given that longer titles/descriptions mechanically increase the pool of potential words to be assigned to a given category. By controlling for the number of publications and for the date of the first publication we attempt to account for the differential breadth of sources consulted and the potential differences in the content depending on the time period of the publication. Finally, we cluster the standard errors at the level of the language family recorded by Berezkin.

Images and Episodes in Folklore Reflecting the Physical Environment

The results are presented in Table 3. Columns (1), (3), (5). and (7) include continental fixed effects and the rest have country-specific constants. In the first two columns the dependent variable is the log count of motifs that mention the word "earthquake." There is a total of 6 such motifs that mention the word earthquake.[17] Typically, these motifs offer a rationale for why earthquakes occur, such as motif $i119$ which is titled: "The dead shake the earth" and has the following description "The earthquakes are produced by the dead who are in the underworld or during the earthquakes the inhabitants of the lower world try to come out." Another motif is $i28a$ titled "Animals produce earthquakes" and the description reads: "Big game animals disappear under the earth and produce earthquakes". Are groups closer to earthquake-prone regions more likely to have such motifs? We construct the distance from the centroid of each group to the nearest high-intensity earthquake region (and follow Bentzen (2015) to define the latter as those in zones 3 and 4). An average group in Berezkin's dataset has 0.10 earthquake-related motifs. However, those located within 100 kilometers of an earthquake zone have on average 0.19 such motifs, approximately 5 times as many compared to groups located outside these areas (mean 0.05). In columns (1) and (2) we show that this pattern is robust to accounting for continental and country fixed effects and the estimated beta coefficients are 0.17 and 0.21, respectively.

In columns (3) and (4) we use as a dependent variable the count of motifs related to thunders. A total of 32 motifs mention this natural phenomenon. To arrive at this number we use "thunder" as a seed in ConceptNet. The related words that are tagged in Berezkin's motifs

[17] We are not using "earthquake" as a seed in ConceptNet because we are specifically interested in earthquakes and not in natural disasters more generally. For example, tagging motifs using the top-related words to "earthquake" from ConceptNet would more than double our count of motifs. This is because we would now include motifs that describe a calamity or a disaster not specific to earthquakes.

are: "thunder," "lightning," "storm," "thunderbolt," and "thunderstorm". Here is an example. Motif, $i2$ is titled: "Lightning from eyes" and reads: "Lightning emerges from eyes or mouth of the being who is thought to produce thunderstorms." This motif is present in 56 societies. Perhaps not surprisingly, most thunder-related motifs are also etiological, i.e., they offer an explanation for the origins of thunderstorms. To measure the lightning intensity of a region, we use the gridded climatologies of the mean lightning flash rates observed by the spaceborne Optical Transient Detector (OTD) and Lightning Imaging Sensor (LIS) instruments from 1995 to 2010 (see Cecil, Buechler, and Blakeslee (2014)). Societies located in regions experiencing intense thunder strikes feature more thunder-related images and episodes in their oral tradition. This is the case both when we compare groups within the same continent and within the same country.

Do societies located in colder environments feature scenes in their oral tradition that depict the cold climates they experience? To answer this question, we use the word "cold" as a seed for ConceptNet. There are 34 motifs that are tagged by the following words: "cold," "cool," "chil," "icy," "warm," "freeze," "frozen," and "frost." Motif $c27$ titled "A horn in the ice" is present in 40 oral traditions and its description reads: "A horned monster breaks ice on river or lake. Usually people walk across frozen body of water, get to see a horn protruding from the ice and try to cut it off. The monster breaks the ice, many people drown".

How does the annual temperature shape the frequency of cold-related motifs? We construct for each society the mean annual temperature between $1900 - 2000$ in a radius of 50km from its centroid. In regions with a mean annual temperature below 10 Celsius, groups' oral tradition have on average 1.44 cold-tagged motifs. This is two and half times larger (0.59) than the corresponding number for groups located in warmer climates. In columns (5) and (6) we show that this pattern holds within continents and within countries.

It is well understood that the environment exerts a significant influence on a group's mode of subsistence. For example, groups residing on more fertile lands are on average more likely to depend on agriculture. In the last 3 columns of Table 3 we ask whether this relationship is also evident in the oral tradition of a group. We use the caloric suitability of agriculture in a given region for crops available before AD 1500 from Galor and Ozak (2015). But how do we get a proxy of the importance of agricultural activities from a society's folklore? We look for the BoW from 3 different seeds which reflect the most commonly cultivated crops worldwide, namely "rice," "wheat" and "corn". Doing so we tag 27 different motifs related to these 3 main crops. The tagged words are: "flour," "oat," "cereal," "grain" "wheat," "maize," "cob," and "corn." Motif $h34g$ is a crop-related motif. It can be found in 35 oral traditions, it is titled: "One grain porridge" and its description reads "One cereal grain (cob, etc.) is enough

to prepare a meal."

An average group in Berezkin has 0.59 crop-related motifs. However, among groups that have 0 such motifs, the median regional caloric suitability for agriculture is $5,277$, whereas the caloric suitability jumps 40% for groups with at least one farming-related motif reaching $7,390$. This pattern is the same when we exploit within-continent variation and within-country variation in columns 7 and 8, respectively. In the last column of Table 3 we add as a control the change in mean caloric suitability from agriculture induced by the post-1500 large-scale population movements. We are doing this exercise in order to indirectly check whether the crop-related motifs in the oral tradition are shaped by the pre-1500 CE agricultural opportunities or by the agricultural options shaped during the last 500 years by the Columbian exchange. The estimated coefficients suggest that although increases in agricultural quality post-1500 enter with a positive sign it is the pre-1500 agricultural landscape that is economically and statistically correlated with crop-related motifs in a group's tradition. Figures $3a - 3d$ depict the relationship between each feature of the physical environment and its intensity in the oral tradition. In particular, the binned scatterplots correspond to the relationships estimated in columns 2, 4, 6, and 8, respectively. Overall, the results in Table 3 provide strong evidence for influences of geography and ecology on folklore-based measures of the natural environment.

4.2 Folklore and the Ethnographic Record

Our goal in this part of the paper is to provide evidence that folklore-based measures of the economy and the polity are in accordance with measures of similar aspects recorded in the EA. Our hope is that by doing so, we can then use folklore to deduce other aspects of a group that are either have incomplete coverage or are plainly absent from the EA, such as the extent of trade. Hence, we view oral tradition and the ethnographic record as providing (noisy, but as shown below, consistent) information for the same underlying social, economic and institutional structure.

To construct a correspondence between a group's oral tradition and its ethnographic record, we linked the societies in Berezkin's database to those in the EA. Out of the $1,264$ groups in the EA, we found a corresponding oral tradition in Berezkin's database for $1,238$ of these societies, resulting in a match rate of roughly 98%. From the 945 groups in Berezkin, we matched 617 to these $1,237$ societies in the EA, implying no ethnographic coverage for approximately one-third of societies for which Berezkin has systematized its folklore.[18] Generally, we run OLS

[18] For these 328 societies, after establishing the link between oral traditions and ethnographic traits, we can use the empirical relationship to reconstruct the missing ethnographic record of these groups.

specifications of the following type:

$$\text{EA Trait}_{i,c} = a_c + \beta \ln(1 + \text{Topic-Specific Motifs}_{i,c}) + \gamma Baseline\ Controls_{i,c} + \varepsilon_{i,c} \qquad (2)$$

where EA $\text{Trait}_{i,c}$ is the trait of interest from the EA for group i located in continent/country c; and $\ln(1 + \text{Topic-Specific-Motifs}_{i,c})$ is the log number of motifs belonging to a given BoW. The term, a_c, captures continent or country-specific constants. Baseline Controls are as described above. See Table 2, Panel B, for the summary statistics of this sample.

Mode of Subsistence in the Oral Tradition and in the Ethnographic Record

Table 4 aims at showing that the folklore-based measures of the specific subsistence modes are in accordance with ethnographic measures of economic activity. Needless to say, we interpret these results as conditional correlations.

We generate the appropriate BoW for each different type of subsistence in the following manner. To capture hunting activities we employ the BoW that results from using "hunt" as a seed in ConceptNet. Doing so, we end up tagging 131 motifs. Words tagged by the resulting BoW include: "hunt," "chase," "deer," "moose," "hunt," "elk," "pursuit," "search," "quest," and "seek." The following hunt-related motif, $h19$, titled: "Raven tries to starve people," is found in 27 oral traditions and its description reads: "Raven hides or scares game animals preventing hunters to kill them. People outwit him." To capture fishing activities we use "fish," as our seed word. This results in 98 motifs tagged as fishing-related ones. The words that are tagged include only "fish," and "fisherman". Motif, $m18b$, is fish-related and it is titled "Getting fishing hook or harpoon" with the following description: "Person turns into a fish to steal fishing hook or harpoon or he turns into fishing hook to catch a fish."

For animal husbandry we use "pastoralism" as our seed in ConceptNet and this results in 27 motifs that are tagged accordingly. The related words in this case include: "cattle," "agriculture," "graze," "herd," "farm," "herdsman," "livestock," and "pasture". Here is the description of a pastoral motif, $k136$: "A lad becomes a master and a leader of great amount of cattle (cows or buffaloes) and meets a princess." Finally, for agriculture we use the BoW explained above.

The dependent variable in Table 4 is the share of subsistence from agriculture in columns (1) and (2), the reliance on herding in columns (3) and (4), and the share of subsistence from fishing in columns (5)-(6) and hunting and gathering in (7)-(8). Each measure of subsistence ranges from 0 to 9, roughly mapping into the deciles of the share of subsistence needs covered by the corresponding activity. The folklore-based measures enter with the expected signs across all specifications when we exploit variation across groups both between and within countries. For example, societies with more crop-related motifs in their oral traditions are systematically

more likely to be farmers, and those with more motifs describing hunting are less likely so. For pastoral societies the beta coefficient on the intensity of pastoral motifs is roughly 0.20.

Predominantly fishing societies have a median of 4 fishing motifs whereas the corresponding statistic for non-fishing groups is 1. Columns (5) and (6) show that this pattern holds when we compare groups within continents and within countries. Finally, hunting-specific motifs are robust features of hunting and gathering societies. The latter are also systematically less likely to have pastoral and farming motifs.

A complementary way to gauge the predictive power of folklore-based measures of subsistence with respect to the shares of subsistence recorded in the EA is to look at the R^2 in specifications without continental or country constants (not reported). The estimates range from 0.10 for the reliance on animal husbandry to 0.16 for the share of subsistence from agriculture, suggesting that a substantial fraction of the observed variation of a given mode of subsistence can be accounted for by variation in the intensity of the subsistence-related motifs. In Figures $4a - 4d$ we show binned scatterplots visualizing the relationships established in columns 2, 4, 6, and 8 of Table 4, respectively.

Uncovering Missing Values of Traits Recorded in the Ethnographic Record

The first two columns of Table 5 follow a strategy similar to Table 4. Specifically, we ask whether folklore-based measures of political complexity predict the strength of political complexity in the EA. The set of words we use to categorize a motif is based on the seed "king" that we input in ConceptNet. Doing so, the following words are tagged: "king," "queen," "prince," "princess," "ruler," and "rein". An average society in the EA has 1.80 king-related motifs. This number jumps to 4.03 for centralized societies and is 4 times smaller (1.04) for non-centralized ones.[19] Columns 1 and 2 in Table 5 show that the same pattern is present within continents and countries. Hierarchical groups during the precolonial era (as recorded by ethnographers in the EA) are systematically more likely to have authority-related motifs. Figure 5a shows the binned scatterplot corresponding to column 2.

Out of the $1,264$ societies in the EA there is missing information on the layers of jurisdictional hierarchy beyond the local community for 135 groups. Can the oral tradition help us recover this information? Figure 5b plots the value of predicted levels of jurisdictional hierarchy based on the intensity of king-related motifs (and the baseline controls of equation (2). To facilitate comparison, besides plotting the average for the groups with missing values, we also plot the predicted values at different levels of observed hierarchy. The average group without a

[19] We follow Michalopoulos and Papaioannou (2013) and classify non-centralized groups as those with 0 or 1 layers of jurisdictional hierarchy above the local community level (variable $v33$ in the EA). See Fortes and Evans-Pritchard (1940) for an original exposition.

record on political complexity is predicted to be roughly at the level of societies organized along petty chiefdoms in the EA. One might expect groups with missing values to be less politically complex. On the contrary, a simple t-test reveals that in fact groups with missing ethnographic records on political centralization are on average more complex based on their oral tradition compared to those groups where such information was recorded. This difference is statistically significant at 5% level. Figure 5*c* depicts the density plots of predicted centralization for those with and without EA-recorded levels of hierarchy, respectively. The latter is shifted overall to the right. This example illustrates the possibility of using a group's oral tradition to recover missing entries of the ethnographic record.

Using Folklore to encode non-extant, Group-Level Characteristics: The Case of Trade

In the last three columns of Table 5, we ask a different question. Motivated by the finding that folklore reflects features of the economic and institutional structure of a group, we use a group's oral tradition to obtain a measure of the intensity of the exchange economy in that society. To do this, we construct a BoW that aims at capturing the presence of trade. Specifically, we use "trade" as the seed in ConceptNet and tag the motifs that contain the related words. These are the tagged words: "exchange," "trade," "swap," "barter," "deal," and "market." The average group in Berezkin's database has 1.27 motifs related to exchange (see Figure 6 for the global distribution of such motifs). Is there a way to verify whether the variation in exchange-related motifs reflects the true underlying trade intensity? Data on the extent of the market economy is not available from the EA, and such estimates are largely missing from the historical record. An indirect way to get at this question is to compare (distance to) historical trade routes to the intensity of trade-related motifs.

In other words, is it the case that groups closer to the preindustrial trade routes are more likely to have trade-related motifs in their oral tradition? To shed light on this question, we used data from Michalopoulos, Naghavi, and Prarolo (2018) that put together for the Old World a comprehensive set of pre-AD 600 trade routes, along with historical harbors and ports before the $5th$ century AD as well as the network of Roman roads, and constructed the distance of each group in Berezkin's database to the nearest pre-AD 600 route. The summary statistics are telling of a robust, broad pattern. Among the 474 societies in Berezkin's catalogue located in the Old World (Africa, Asia, Europe), those within 100 kilometers of ancient trade routes have an average of 3.52 exchange-related motifs, a number twice as large compared to groups located farther away (which have 1.78 such motifs).

Columns 3 and 4 of Table 5 show that this pattern is not driven by broad group differences

across continents or modern-day countries, highlighting the usefulness of folklore in quantifying missing yet important aspects of preindustrial societies. In column 5 we add a control reflecting how much the distance to trade routes decreased in between 600 AD and 1800 AD. Interestingly, the latter enters insignificant, whereas the pre-AD 600 coefficient remains precisely estimated. This pattern echoes what we found earlier regarding the role of pre-1500 and post-1500 land quality for agriculture and crop-related motifs, and suggests that although we only have a snapshot of the oral traditions across societies in late 19th century, elements of folklore are likely to encode information of the economy and the polity harking back several hundreds of years.[20]

4.3 Folklore and Historical Norms

So far, we have shown how the oral tradition of a given pre-industrial society may complement our reliance on the EA, both deepening and broadening our understanding of a group's economic, social and institutional background. Besides the value of having two (noisy) sources to reconstruct societal historical attributes, below we show how we use folklore to reconstruct the historical beliefs, attitudes, and norms via text analysis. Naturally, in absence of alternative proxies of cultural norms we cannot directly check how accurate are the values elicited from a group's folklore. Nevertheless, by establishing that the content of folklore is broadly consistent with the known ethnographic material it increases our confidence that the historical values inferred from the oral tradition may be useful proxies of the unobserved historical cultural norms.

After discussing the methodological framework, we attempt an empirical assessment of famous conjectures among social scientists. Specifically, through the lenses of folklore we investigate the original affluent society hypothesis, the culture of honor among pastoralists, the importance of family in groups formed along extended family lines, and the relationship between statehood and rule-following norms as well as trade. This list of conjectures is by no means exhaustive. Our choice is mainly driven by the fact that besides their outsized importance in the social sciences, these hypotheses have been the departing point for a series of studies among economists interested in how culture shapes contemporary outcomes.

A Methodological Digression

The set of values that one may extract from folklore is potentially very large. How do we discipline our choices? Broadly speaking, there are two categories of questions one may ask.

[20] In the LIWC dictionary there is an entry reflecting concern with money. This is a broader category compared to the one on trade from ConceptNet. In Appendix Table 2 we show that a similar pattern obtains when we use the motifs tagged by the money-related BoW or the first principal component between the two BoW.

On the one hand, we could test a specific hypothesis. For example, below we will show how a group's oral tradition can shed light on several conjectures from anthropology linking a society's historical mode of subsistence or type of social organization to a given psychological disposition.

On the other hand, one may prefer to let the data "tell" whether there are particular values that stand out from a group's folklore. This is a case of feature selection, i.e., the process of selecting a subset of relevant variables for use in the model. In the era of big data there are numerous methods that have been proposed to help with feature selection. In the context of linear regression models and in the presence of many correlated predictor variables, the general solution is to reduce variance at the cost of introducing some bias. This approach is called regularization and helps the predictive performance of the model. There are various shrinkage techniques including Lasso, Ridge, and Elastic Net (a convex combination of the previous two).[21]

To address either inquiry (test a specific hypothesis, or perform feature selection) one needs to define what is the underlying set of potential psychological states. This question has no clear answer. For example, ConceptNet, a great resource for obtaining related words for virtually *any* seed (word), encompasses much more than psychological states, personal concerns or moral values. This is where existing psychosocial dictionaries can help. Specifically, we discipline our choice of BoW using the entries from two widely-used psychosocial dictionaries that map psychologically cross-validated BoW into specific values and states of mind. See the discussion in Section 3. Using the entries from the LIWC and the MFD dictionaries has clear advantages by delineating the pool of potential values. Nevertheless, an important caveat to keep in mind is that all BoW are based on contemporary English. Although translation issues are less pressing since Berezkin has done the translation of the motifs himself, contemporary words that map into personal concerns and psychological states may not necessarily have the same meaning in the historical context that tales and myths describe.[22]

Sometimes, a given BoW from LIWC/MFD closely reflects what we want to capture; other times, however, a given entry may be only loosely related. In these cases, we will use specific BoW selected from ConceptNet and combine related concepts from LIWC/MFD via a principal component analysis.

[21] A thorough investigation of what the alternative algorithms deliver in the context of oral traditions is beyond the realm of the present study. Nevertheless, below we will sometimes comment on the BoW chosen by the rigorous lasso. The latter was introduced by Belloni et al. (2012) and is a data-driven method of choosing the penalty level that delivers asymptotically valid estimates under non-Gaussian, heteroscedastic disturbances.

[22] Ideally, we would like to have the BoW from LIWC, MFD and ConceptNet to be based and cross-validated on the historical context of the time period an oral tradition corresponds to.

Testing Specific Conjectures: The Case of Correlated Concepts While testing a specific hypothesis, it quickly becomes clear that the uncovered correlations may suffer from omitted variables bias. Suppose that one researcher finds a correlation between psychological state X and group attribute Y. One may naturally wonder whether this is a robust association or whether there is another psychological trait Z, that is correlated both with X and Y resulting in a spurious correlation between the latter two. This is a case of correlated concepts. How best to deal with this? Usually, one would leverage prior knowledge of co-occurrence of psychological/moral traits and control for them appropriately. However, in this case we have weak priors about which BoW are correlated and, moreover, there are dozens of potential BoW from LIWC and MFD. In other words, we do not know what is the "right" set of controls among the numerous BoW.

To address this issue we take advantage of recent developments in estimating structural parameters in linear, sparse, high-dimensional models with many controls, which is our case. Specifically, we use the "post-double-selection" methodology (PDS) proposed by Belloni et al. (2012, 2014, 2015). The idea is to separately lasso over the pool of potential BoW on the dependent variable, Y, and on the BoW of interest, X, and include the union of variables selected as controls in the OLS regression of Y on X . The intuition is straightforward. Find the variables systematically related to either Y or X and explicitly account for them when estimating the association between the two.

Note that the discussion above does not touch upon the issue of causality, i.e., whether X causes Y or the other way around. Nevertheless, it provides an informed way to discipline the choice of potential control variables. Below we will be showing results both with and without the PDS-selected covariates.

4.3.1 Foraging Societies

The Original Affluent Society Hypothesis Before Sahlins (1972), anthropologists portrayed the forager's lifestyle as an indigent one, with day-long toiling to obtain the means necessary to survive and coping with a marginal environment, leaving little if any time for leisure. In 1972 Marshall Sahlins offered a drastically different take on this. Drawing on data from a small number of foraging societies, he argued that hunter-gatherers were able to meet their needs by working roughly $15-20$ hours per week, significantly less than the hours worked by industrial workers, concluding that with economic development, the amount of work actually increases and the amount of leisure decreases. This radically different view espoused by Sahlins has become popular among anthropologists but has also generated vigorous criticism, see Kaplan (2000). To settle this open question in a definitive manner one would need time-use

data for the universe of hunting and gathering societies. An alternative route is to take advantage of a society's folklore. To the extent that a group's lifestyle is imprinted in the images and episodes of its oral tradition then one might expect to find more leisure-related motifs in societies where leisure activities are more common.

The LIWC dictionary, among the BoW reflecting personal concerns, includes an entry on "leisure" that is useful in our context. Leisure–related motifs that are tagged by the corresponding LIWC category include words like: "celebrate", "dance", "entertain", "dream", "fun", "game", "joke", "sing", "play", "art", and "relax." Here are a couple of motifs prevalent among foraging societies in North America; "Person joins dancers but then understands that these are trees or reeds moved by the wind". "Person plays throwing his eyes or his tooth up or away. Eyes or tooth first come back to eye sockets or mouth but eventually are lost." There are 175 societies in the EA that derive their livelihood predominantly from either hunting or gathering. The median foraging society has 17 motifs that are classified as related to leisure whereas non-foraging ones have only 10 such motifs.

Table 6 has the following structure, which will be repeated when we explore the association between an observable economic/institutional trait and a given concept. The first two columns exploit within-continent variation whereas columns (3) and (4) compare the oral traditions of groups located in the same country. Figure 7*a* shows the binned scatterplot of the relationship estimated in Column 1 of Table 6. All columns include the baseline controls explained in equation (2). The *additional controls* in the even-numbered columns are the BoW selected by the post-double lasso selection method over all entries in the LIWC and MFD dictionaries. See discussion above. Doing so, in column (2) ((4)), for example, we control for the frequency of motifs reflecting an additional 6 (8) concepts/BoW that are found to be highly correlated in the first stage. To fix ideas "leisure"-related motifs systematically co-occur with motifs on "motion". Hence, the frequency of motion-related motifs in a group's oral tradition is part of the additional covariates.[23] Comparing the coefficient estimates without and with the additional controls is instructive of how much of the covariance between the BoW of interest and the corresponding trait is driven by concepts that are correlated both with the RHS and the LHS variables. Accounting for the lasso-selected BoW in columns (2) and (4) does not change the estimated relationship either qualitatively or quantitatively. Hunting and gathering societies feature in their oral tradition themes systematically related to leisure.[24]

[23] Although we do not report the variables selected by the first stage of the post-double lasso selection regressions, these chosen covariates offer a unique glimpse into the covariance structure of psychological states and personal concerns in a group's oral tradition.

[24] We find a similar pattern but with the opposite sign when we focus on "work"-related motifs according to the LIWC (not reported for brevity). Foraging groups have on average 22 such motifs compared to non-foragers that have on average 26 such images. Note that this is not a mechanical relationship. It is not the case that when a group features more leisure-tagged motifs it has fewer work-related ones. The empirical relationship

Although one needs to keep in mind that leisure-related words are classified using a contemporary dictionary, these robust associations tentatively offer large-scale evidence in support of Sahlins (1972) thesis. Comparing oral traditions across societies at different stages of development, there is a gradient in the intensity of leisure-related images; those images decrease as societies transit from hunting and gathering to agriculture. Predominantly agricultural groups show the opposite to foraging communities' pattern. Leisure-related motifs are relatively scarce and work-related ones abound.[25]

4.3.2 Pastoral Societies

The Culture of Honor Anthropologist Walter Goldschmidt was perhaps the first to observe that among African cultures "pastoralists would do more 'acting out' in their interpersonal relationships, being more ready to express anger and to take direct action, while the farmers would suppress their negative emotions and restrain their action" (1971, 16–17). Just as pastoral environments are thought to favor a psychological phenotype that is sensitive to honor and prone to aggression, it is similarly thought that horticultural environments select for the opposite psychological profile.

Edgerton's seminal work in 1971 offers a fascinating examination of the psychological profile of farmers and herders across four groups in East Africa. Conducting an extensive comparative psychological study he concluded that herder attitudes were associated with a number of attributes including direct aggression, independence, fear, bravery, and brutality. Bolton et al (1976), focusing on two Peruvian villages of farmers and herders, found that these differences were manifested already among children of 5 to 7 years old, suggesting the formation of these psychological traits at a very young age.

Motivated by these works Nisbett and Cohen (1996) conjectured that the high prevalence of homicides in the South of the United States can be attributed to its initial settlement by herders from the fringes of Britain in the 18th century, who transplanted a culture of honor and concomitant aggression. Grosjean (2014) provides robust evidence along these lines. The idea behind this link is that pastoral societies are likely to rely heavily on aggression as a way to deter theft from their herd. In such environments of imperfect property rights and easily movable property, creating a reputation of readiness to exact revenge may prevent theft.

between these BoW is if anything weakly positive.

[25] Setting aside Sahlin's influential conjecture, one might be interested in knowing what features are conspicuously present (or absent) from the oral tradition of a typical foraging society. We provide a partial answer to this question by investigating the features selected by the feasible Lasso method proposed by Belloni et al (2012). When we cluster the standard errors along the language family level (variable v98 in the EA) and partial out the baseline controls, the feasible Lasso selects 4 BoW as predictive of predominantly hunting and gathering societies. Overall, foragers' oral tradition abounds with "leisure"-related motifs whereas images and episodes related to "money", "religion" and "home" are very much predictive of non-foraging cultures.

A natural question is whether themes of aggression are more frequent in the oral tradition of pastoral groups. Although LIWC does not have an entry focusing clearly on aggression and retaliation it does feature a BoW reflecting "anger" among the list of psychological states regarding negative emotions. Examples of words in this BoW are: "aggressive", "angry", "revenge", "assault", "brutal", "cruel", "kill", etc. We complement this entry with a BoW from Conceptnet on "retaliation" to obtain a more precise measure of motifs that are directly related to instances of retaliation. The words tagged in the motifs are: "revenge", "avenge", "retaliate", "wrath". Predominantly pastoral groups have on average 41.68 (1.80) motifs reflecting anger (retaliation), whereas the corresponding number among non-pastoral societies is 26.86 (1.17).

To quantify the intensity of violent motifs in a group's oral tradition, we take the first principal component of "anger" and "retaliation" BoW.[26] Columns 1 to 4 in Table 7 Panel *A* show that there is a robust association between the frequency of retaliatory, anger-related images in a group's oral tradition and its mode of subsistence. Figure 7*b* displays the binned scatterplot of the relationship estimated in Column 1. Pastoral groups have disproportionately more violent motifs compared to others. This relationship obtains both when we exploit within continent and within-country variation and when we account for correlated BoW with the post-double lasso selection method. When we do so, the coefficient estimate of interest drops moderately, which implies that there are other BoW in LIWC and MFD that are correlated both with the intensity of anger in a group's folklore and are more common in herding societies. For example, it appears that anger-related BoW are positively correlated with BoW associated with authority/subversion in the MFD dictionary. Motifs tagged as reflecting deference to authority are also more common across pastoral societies. Hence, once we control for the share of motifs reflecting deference to authority in the second stage of the post-double lasso selection method, the association between pastoral groups and anger-related motifs weakens.

Establishing that pastoral oral traditions feature more motifs depicting anger and retaliation is important for two reasons. First, it provides large-scale evidence in favor of an influential (and intensely debated) conjecture about the culture of honor and aggression among pastoral societies. Second, this finding also helps us understand how a particular attitude survives across generations even when the original conditions, that made this type of cultural adaptation emerge, no longer apply. A group's collective memory enshrined in its oral tradition may be the vehicle of cultural transmission.

[26] The first principal component explains 83% of the common variance of these two measures of aggression across oral traditions. We focus on the first principal component because its eigenvalue is 1.67 (one being the rule of thumb), while the eigenvalue of the second principal component is strictly less than one. In Appendix Table 3 we show that the results are similar when we use as the dependent variable each BoW, separately.

4.3.3 Societies Organized Along Extended Family Lines

Family Members in the Oral Tradition Numerous studies have documented the importance of family ties for understanding a variety of policies and contemporary attitudes related to trust towards family members and demand for regulation of labor markets, among others. See Alesina and Giuliano (2013), Alesina, Algan, Cahuc, and Giuliano (2015), Enke (2018), among others.

Here we explore whether groups organized along extended family lines are more likely to feature family members in the images and episodes of their oral traditions. This question is motivated by the fact that unlike nuclear families, in extended-family societies socialization is more likely to take place within the confines of the extended family network. If this is the case, then one would expect to see family members often being the protagonists in the stories and tales being passed from one generation to the next. This is a straightforward exercise to conduct since among the LIWC entries that describe social processes, there is a BoW specific to "family". Among the $1,237$ groups recorded in the Ethnographic Atlas, 361 are coded as being domestically organized along independent nuclear families with the remaining 876 featuring polygynous, and extended families. In the latter, family members show up in 22% of the motifs whereas the corresponding statistic is 19.9% for societies with a nuclear structure. This simple difference in means is statistically significant at 1% level.

In Table 7− Panel B we explore the robustness of this association. Specifically, in columns (1) and (2) we look within continents whereas in columns (3) and (4) we add country-specific constants. Figure 7c displays the binned scatterplot of the relationship estimated in Column 1. The odd-numbered columns include only the baseline controls whereas in the even-numbered ones we add the controls selected by the post-double lasso selection method. Adding these additional BoW weakens the relationship between extended family groups and the prevalence of family-related motifs but it remains both economically and statistically significant. In the most restrictive specification of Table 7 Panel B, nuclear-family groups have 5% fewer motifs where family members are mentioned.

This finding matters not only for our understanding of how the content of motifs reflects the mode of social organization (in this case nuclear versus extended families) but, as will become apparent, this differential representation of family members in the narratives of groups is an important predictor of how much individuals trust their own kin today.

4.3.4 Political Centralization

Historical states have fascinated economists with an accompanying increase in the studies exploring their legacy and origins over the last two decades. Thanks to the thorough historical

work and the wealth of ethnographic studies available, our understanding of societies with complex political legacies is substantial. Nevertheless, there is one dimension for which we know relatively little. This regards the extent of trade whose link with statehood has been often assumed but comprehensive evidence is lacking. For example, Robert Bates has famously linked the roots of statehood to the gains that can be reaped from the promotion of markets. In regions where trade could flourish more easily, states would arise to facilitate production and exchange. Despite leveraging interesting geographic variation in trade potential, see Fenske (2014), there is an inherent difficulty in testing any hypothesis involving the role of trade in history. This is particularly true for societies outside Europe and China where information on the importance of exchange is at best limited.

The Importance of Trade In this section we explore the relationship between statehood and the intensity of trade motifs in a group's folklore. The measure of trade intensity we have already constructed takes into account only trade-related words from ConceptNet. Nevertheless, in LIWC among the BoW on personal concerns there is an entry on "money" which encompasses trade-related words. Hence, we combine the trade-related motifs from ConceptNet with the money-related ones from LIWC to obtain a set of motifs that are related to trade and money more broadly. Specifically, we take the first principal component that explains 90% of the common variance of the two measures across the 945 oral traditions. We focus on the first principal component because it loads positively on both variables and has an eigenvalue of 1.81, while the eigenvalue of the second principal component is 0.2. In Appendix Table 4 we show that the results are similar when we use as a measure of exchange each BoW separately.

Armed with the measure on the intensity of exchange in the oral tradition, in Panel *A* of Table 8 we explore its relationship with the degree of political centralization. Again, we make no attempt to get at the question of what causes what; our goal is to simply offer illustrative correlations of a pattern that has been much theorized upon with few empirical counterparts.

Among stateless societies, i.e., those without any levels of jurisdictional hierarchy above the village level, motifs on trade and money are notably absent, with a mean of 0.72 and 3.94, respectively. On the contrary, among groups organized as states the relevant statistics are roughly 5 times larger (3.6 and 19.83, respectively). Table 8-Panel *A* suggests that this pattern remains strong both when we explore within continent and within country variation in columns (1)-(3) and (4)-(6), respectively, as well as when we additionally account for the BoW from LIWC and MFD identified by the post-double lasso selection method in specifications (2), (3), (5) and (6). Figure 7*d* displays the binned scatterplot of the relationship estimated in Column 1.

In columns (3) and (6) we control for the shares of subsistence that come from agriculture and from pastoralism. Two points are worth discussing. First, political centralization remains a robust correlate of trade and money beyond the relationship between the mode of subsidence and exchange itself. Second, the intensity of money-related motifs for agricultural groups is no different compared to that of foraging groups, whereas pastoral groups are systematically more likely to feature exchange-related themes in their oral tradition. This pattern is consistent with the observation that farming communities in the preindustrial era were not necessarily more likely to engage in trade—agricultural groups could satisfy a large part of their subsistence needs from agricultural products, whereas pastoral ones, given the limited set of resources they produced, had to rely more on trade and exchange for their survival. Richerson, Mulder, and Vila (2001), for example, observe that "despite the emphasis on animals, most herders are dependent on crop staples for part of their caloric intake ... procured by client agricultural families that are often part of the society and the presence of specialized tradesmen that organize the exchange of agricultural products for animal products."

Rule-Following Norms Some studies have identified robust associations between the historical statehood and contemporary economic performance at the country and ethnic homeland level, see Bockstette, Chanda, and Putterman (2002) and Michalopoulos and Papaioannou (2013), respectively. Others have looked at how the experience of centralized forms of governance influences the norms towards rule following today. From a theoretical standpoint, the answer for the latter is ex ante ambiguous, and it boils down to whether institutions crowd in or crowd out rule-following behavior. From an empirical point of view, the existing findings go in both directions. For example, Lowes, Nunn, Robinson, and Weigel (2017) compare descendants of a centralized group, the Kuba, to those of neighboring stateless groups in Congo DRC and find in field experiments that the former are less likely to follow the rules and more likely to cheat, suggesting a substitutability between the strength of the state and rule-following norms. On the contrary, Dell, Lane, and Querubin (2017) focus on the Dai Viet-Khmer boundary within Vietnam, and comparing adjacent villages, find that a strong historical state crowded in village-level collective action and quality of local governance. Can the oral tradition of a group help to shed light on this question?

The MFD dictionary has 82 words that are meant to capture virtues and vices of leadership and followership, including deference to legitimate authority and respect for traditions.[27] Among those, the most common words in the tagged motifs are: "order," "control," "status,"

[27] We combine both BoW on vice and virtue for authority constructing a BoW on authority overall. From the 287 motifs tagged by the MFD as authority-related motifs 278 are tagged by the virtue-related BoW and only 9 as vice-related.

"serve," "leader," "respect," and "hierarchy". Here is a description of such a motif present in 26 societies: "For finding the lost husband (wife, savior), person sets up an inn (bakery, bathhouse) where guests are served gratis. This attracts people. One of the guests proves to be the lost one or the guest tells a story that helps to find the lost one." Another one reads: "Person who gives difficult tasks to the hero is a prominent figure in social hierarchy, i.e., a head of political unit of community- or higher level and not a mythical being." This motif is present in 228 societies in Berezkin's dataset.

We complement the entry on authority/subversion from the MFD dictionary with the BoW from ConceptNet on "respect". We do so because the MFD BoW on authority/subversion is based on a broad set of words whereas the ConceptNet BoW for respect has significantly fewer. As a result, 287 motifs are tagged by the authority/subversion BoW from MFD whereas only 26 motifs are tagged by the "respect" entry in ConceptNet. The words tagged by the latter are: respect, admire, affect, gratitude, honor, regard, and sympathize. Below, we use as the dependent variable the first principal component of the two.[28]

A simple cross-tabulation of the intensity of these categories and state centralization as recorded in the EA is indicative of the underlying pattern. On the one hand, stateless groups have on average just 0.45 motifs that are tagged as respect-related by ConceptNet (the respective number for the authority/subversion entry from MFD is 6.99). On the other hand, the corresponding number for centralized states is four times as large with an average of 2.34 (22.19). In Table 8, Panel B, we show that this pattern is robust to adding continental fixed effects in columns (1) and (3), and comparing groups within the same modern-day-country boundaries in columns (2), (4), (5) and (6). Moreover, controlling for additional BoW chosen by the post-double selection method, the coefficient estimates of interest decline in magnitude but remain precisely estimated. Finally, in the last two columns we show that this correlation is present both in groups that have high gods and in groups without high gods. Across these different samples and conditioning sets we find that political complexity goes in tandem with respect (for authority) motifs. Figure 7e shows the binned scatterplots of the relationship estimated in Column 1.

All in all, the results in Table 8, Panel B, offer large-scale support in favor of arguments put forth that predict a complementarity between statehood and rule-following behavior, see (Weber (1976) and Foucault (1995) among others).

[28] In Appendix Table 5 we show that results are similar using each BoW separately.

4.4 Historical Norms and Contemporary Beliefs and Attitudes

Do historical norms correspond to contemporary attitudes? Folklore is believed to be the "intellectual remains of earlier cultures surfacing in the traditions of peasant class". If so, to what degree, do cultural values encapsulated in the oral tradition predict values and beliefs today?

To answer this question, we turn to the World Value Surveys (and its European equivalent—the European Value Surveys). The first step is to assign an oral tradition to each respondent based on her information regarding ethnicity, language spoken at home, and language spoken at interview. Out of the 417,347 respondents, for which at least one of the three characteristics is present, we have recovered an oral tradition for 368,951 individuals. The procedure is the following: whenever information on ethnicity is available, a respondent is matched to the folklore tradition(s) of his ethnic identity. If ethnicity is missing or unknown, we look at the variable indicating the language the respondent speaks at home and assign the corresponding oral tradition.[29] When no oral tradition could be matched from the first two steps, we use the language a respondent speaks at the interview.[30] This accounts for about 25%, or 86,246 respondents, of the sample. For regions in Europe with a well-known regional identity whose folklore is also present in Berezkin's catalogue, we institute an overriding rule: we assign all respondents sampled in that specific region to the regional folklore tradition. This is the case for the following regions: Scotland, Ireland, Wales, Aragon, Sicily, Sardinia, Eastern Sami, Western Sami, Gagauzia and Kashubia. Behind the links we have constructed lies a deeper question, i.e., what is the vehicle by which an oral tradition is passed from one generation to another? Is it one's ethnicity? Is it the language one speaks at home, the language one uses in his communications outside home, his region or a combination of all of the above? Our current matching prioritizes ethnicity over home language and the latter over the language of the interview.[31]

Having constructed a correspondence between an individual and her oral tradition, we

[29] It is possible for one language/ethnicity to correspond to several oral traditions. When this the case (like for the "indigenous" in Guatemala), we assign to these groups the simple average of folklore measures of the indigenous societies present in Guatemala (in Berezkin's corpus). Other times, a respondent answers with a country as his/her ethnic identifier. In this case, we created country-level BoW. We did this in the following manner. First, we matched groups in the Georeferencing of Ethnic Groups Dataset (GREG) to an oral tradition. We then generated for each ethnic group in GREG the corresponding BoW. Finally, to arrive at a country-level BoW we took the weighted average of the BoW of the country's groups using as weights the groups' population shares in the 1990's.

[30] When a respondent speaks English in the interview in a former British colony but he speaks a different language at home which we cannot match to an oral tradition, we err on the conservative side and do not assign to her the English oral tradition.

[31] An alternative route to reconstructing an individual's oral tradition would be to use some kind of maximum likelihood estimation minimizing the distance between folklore-based values (of one's ethnicity, language and region) and current attitudes. Doing so, one could uncover the closest oral tradition.

attach to the respondents their historical values based on their folklore.

The empirical specifications we use in this part of the study take the following form:

$$\text{Attitude}_{i,g,c} = a_c + \beta \text{Historic Value}_g + \gamma \text{Baseline Controls}_g + \zeta \mathbf{X}_{i,g,c} + \varepsilon_{i,g,c}, \quad (3)$$

where $\text{Attitude}_{i,g,c}$ is the answer given by individual i, with oral tradition g, residing in country c. $\mathbf{X}_{i,g,c}$ is a vector of individual characteristics including age, age square, sex, 9 educational attainment indicators and 50 religious denomination fixed effects. Standard errors are clustered at the oral tradition level and a_c reflect country-of-residence specific constants. The baseline controls are the same as those discussed above.[32]

We use the BoW we constructed in the previous section to capture historic values from an oral tradition, $Historic\ Value_g$.

Respect (for Authority) in a Group's Folklore and Rule-Following Today We first examine how the intensity of "respect"-related motifs in a society's folklore influences tolerance of antisocial behavior today. The reason we look at these values is because there is a vibrant literature that attempts to understand the variation in contemporary rule-following norms across countries. Starting with the well-identified study by Fisman and Miguel (2007) who show that cultural norms towards corruption are systematic determinants of rule-following behavior, many studies try to shed light on how these rule-abiding norms rise in the first place and get transmitted over time.

Given the interest of the academic and policy-making community in this aspect of culture it is not surprising that the WVS-EVS has several questions that get at different instances of condoning socially-deviant behavior. In Table 9 we use 4 of these questions, $F114 - F117$. Specifically, in column 1 the dependent variable reflects how justifiable the respondent finds claiming government benefits to which she is not entitled. In column 2 (3) the variable of interest is the extent to which a respondent justifies avoiding paying the fare on public transport (cheating on taxes) and finally in column 4 the question gauges how comfortable is the individual with someone accepting a bribe. These 4 questions are on the same scale and range from 1 to 10 where 1 implies that the respondent finds this specific action "never justifiable" and 10 "always justifiable". So, higher values indicate greater tolerance for rule-breaking behavior.

The first 4 columns of Table 9 Panel A paint a clear picture. Respondents belonging to groups whose oral traditions have more images and episodes on respect for tradition are systematically more likely to condone instances of anti-social behavior, like exploiting government benefits, cheating on taxes, avoiding transport fare or accepting a bribe. The estimated beta

[32] We focus on oral traditions with at least 40 respondents to make sure that our results are not driven by leveraging variation from groups with a handful of respondents.

coefficients are also sizeable. A one-standard-deviation increase in the intensity of "respect"-related motifs decreases tolerance for socially deviant behavior by around 0.7 standard deviations. We obtain the same pattern at the group level (Panel A) and at the individual level in Panel B. The individual-level specifications are interesting because they show that the pattern uncovered across oral traditions (i) is not driven by cross-country specific differences in folklore and (ii) remains robust when we exploit across-individuals differences in education, age and religious denomination. Within major and minor religious groups, variation in the frequency of motifs related to authority/subversion systematically predicts low tolerance towards rule-breaking behavior.

The first 8 columns of Appendix Table 6 in Panels A and B show that the pattern is qualitatively similar if instead of the first principal component we use the individual BoW on "respect" from ConceptNet and "authority/subversion" from MFD, respectively.

Family-Related Motifs and Trust in Family Does exposure to stories at an impressionable age, where family members are the main characters, increases one's trust to family members? Question $D001_B$ in the WVS-EVS asks the respondent how much he/she trusts her own family and the answers range from 1 to 4, where 1 indicates complete trust (this is the case for 83% of the respondents) and 4 complete distrust (this is the case only for 0.65% of the sample). Here is a rather common family-related motif present in 74 societies: "Many brothers marry or have to marry in such a way that all their wives are sisters."

A one-standard deviation increase in the frequency of motifs featuring family members increases trust in family by 1.02 standard deviations, suggesting a quantitatively important association (column 5 in Panel A of Table 9). Column 5 in Panel B shows this to be true also when we explore individual-level variation. In the sixth column of Panels A and B we ask whether this link masks an increase in generalized trust among those exposed to an oral tradition rich in episodes featuring family members. Specifically, we control for the level of generalized trust (variable $A165$) at the group and individual level, respectively. The association between the frequency of family-related motifs and trust towards family members remains intact.

Leisure-Related Motifs and Importance of Enjoying one's Time Above we established that leisure-related images are more common among hunting and gathering societies. We interpreted this finding as evidence in favor of a lifestyle of relatively abundant leisure for these groups. Naturally, one may wonder if such motifs are consequential for contemporary norms, i.e., when a society's myths and legends feature images of fun and amusement, do members of such groups place more importance on having a good time? Question $A192$ gets at

this point. In particular, the question reads: "It is important to this person to have a good time" with potential answers ranging from 0 ("very much like me") to 6 ("not at all like me.") Hence, higher values indicate disagreement with the statement in question. Answers to this question are fairly uniformly distributed across the different dispositions. Column 7 reveals a robust inverse relationship between the frequency of leisure-related motifs in an oral tradition and the extent of disagreement with the importance of having a good time. The relationship is also economically meaningful with an estimated beta coefficient of 0.6. Is this pattern driven by country-specific features, or individual differences in age, education, gender and religious identity? Column 7 in Panel B shows that this not the case.

A similar patterns emerges when we look at individual attitudes towards non-work in column 8. Specifically, question $C038$ reads: "People who don´t work turn lazy" and higher numbers suggest that the respondent increasingly disagrees with this statement. People whose oral traditions have more motifs related to leisure are systematically less likely to think that people who do not work turn lazy. Exploiting within-country, cross-oral tradition variation delivers the same pattern. Overall, the evidence in columns 7 and 8 in Panels A and B of Table 9 suggests that exposure to images and episodes of leisure seems to make people value having a good time, and be less likely to think of someone as lazy if she is not working.

Anger/Retaliation-Related Motifs and Attitudes Towards Violence Today Attitudes toward violence vary significantly across groups. For example, in the question whether "it is justifiable to use violence against other people", $F114_03$, there are groups like the Tswana, on the one end, whose respondents are very much likely to find such an action justifiable compared to the Poles, on the other end, who almost unanimously reject the use of violence.[33] 5.5% (51%) of Tswana's motifs feature an act of retaliation (anger) compared to just 0.6% (34%) for the images in the oral tradition of the Poles. Is this pattern more general? That is, can part of the variation in attitudes towards violence be attributed to the intensity of anger and retaliation-related motifs in the groups' oral tradition? In Column 9 the dependent variable reflects the share of people within a group that find violence more justifiable than not. To capture the psychological proclivity of groups toward violence we use the first principal component of motifs tagged as related to anger (by the LIWC) and retaliatory ones as tagged by ConceptNet. Groups whose folklore displays images of anger and revenge are more accepting towards violence overall. Panel B shows that this pattern is the same at the individual level within countries.[34]

[33] Tolerance towards violence ranges from 1 which is never justifiable to 10 which is always justifiable. So, values above 5 indicate a respondent who is more likely to justify the use of violence than not.

[34] In columns 9 and 10 in Appendix Table 6 we use the BoWs related to anger and retaliation separately. It is the retaliatory motifs that seem to have all the explanatory power vis a vis current attitudes towards violence.

Our initial examination of the WVS-EVS with the psychological states and moral values captured by BoW from the LIWC and MFD dictionaries indicates a striking consistency between past values derived from folklore, and contemporary values reflected in the WVS-EVS regarding rule-following, trust towards family members, acceptance of violence and attitudes towards enjoying one's life. Images and episodes encapsulated in folklore related to these features appear to have a lasting life and more suggestively, but quite possibly, are still shaping the way individuals perceive the world. This reveals the potential for folklore to be an anchor for past values and to serve as an important benchmark in the research of culture.

5 Concluding Remarks

The economics literature on the cultural determinants of growth, starting with Landes (1998) , have placed a great emphasis on culture for determining contemporary economic and political outcomes. To overcome the issue of endogeneity, promising instrumental variable strategies have been devised linking historical accidents and geographic endowments to contemporary beliefs and attitudes, often looking at individuals no longer living in their ancestral homelands (following the epidemiological approach of Fernandez (2011)). Nevertheless, there has been a missing element in this literature. Namely, the absence of historical proxies of beliefs has severely limited the debate about the transmission of culture.[35] In this study we propose a way to close this gap by integrating folklore into our toolset.

Specifically, we do four things. First, we introduce and describe a novel catalogue of oral tradition across approximately $1,000$ preindustrial societies assembled by the eminent anthropologist and folklorist Yuri Berezkin. Second, following a dictionary-based method, we quantify several aspects of folklore related to the physical environment, the mode of subsistence, and its institutional complexity. We show that these folklore-based measures are predictive of the observed natural landscape, and of the economic and societal structure recorded in the *Ethnographic Atlas*. We then describe how one may utilize a society's oral tradition to quantify unknown economic and institutional aspects including the importance of trade and the degree of jurisdictional hierarchy for those groups where such information is missing.

Third, motivated by Bascom's (1953) view of folklore as a depository of a group's beliefs and attitudes, we make a first attempt to uncover these norms by applying a dictionary-based method to the motifs of the recorded oral traditions. We use the LIWC and the MFD dictionaries, which have been widely applied in linguistics, psychology, sociology, and anthropology to analyze text. Armed with these historical values we proceed in two steps. First, we use them

[35] See Chen (2013) and Galor, Ozak, and Sarid (2016) for attempts to link linguistic features regarding, for example, the structure of the future tense to cultural attributes today.

to assess the empirical content of an array of influential conjectures among anthropologists regarding the original affluent society, the culture of honor among pastoralists, the importance of family for groups organized along extended-family lines, and the relationship between statehood and trade and rule-following norms, respectively. Finally, we explore how the norms deduced from a group's oral tradition are predictive of contemporary beliefs and attitudes.

We view this study as a springboard for further research. For example, one can utilize folklore to derive bilateral measures of cultural proximity, see Spolaore and Wacziarg (2009), or explore how different geographical traits influence the content of oral traditions. For example, what do groups located in malaria-prone regions, fertile territories, or rugged terrains, "talk" about? Moreover, folklore may shed light on which values are largely stable and which ones are subject to change. An alternative dimension along which folklore can be employed is related to Berezkin's work, that is, using it to trace the groups' historical migration paths. Also, although obtaining time variation in folklore is challenging given the inherent uncertainty with timing the origin of a given motif, one may extend this analysis to construct the psychological content of contemporary popular culture as reflected in song genres, for example.

Finally, one could investigate whether the content of children's books read today in different parts of the world (or the US) is similar to the oral tradition of the respective places/groups. We posit that the degree of continuity in the narratives between contemporary children books and the folktales and myths of the respective societies is a direct measure of the rate at which ancestral norms are intergenerationally transmitted. Given the versatility of folklore as a vehicle of obtaining a unique (and perhaps our only) view of our ancestral oral cultural heritage, we expect it to be widely used among scholars interested in the historical origins of comparative development, culture and beyond.

References

Alesina, A., Y. Algan, P. Cahuc, and P. Giuliano (2015): "Family Values and the Regulation of Labor," *Journal of the European Economic Association*, 13(3), 599–630.

Alesina, A., and P. Giuliano (2013): "Family Ties," in *Handbook of Economic Growth, Volume 2*, ed. by P. Aghion, and S. Durlauf, vol. 2. Elsevier, Amsterdam, Netherlands.

Alesina, A. F., and P. Giuliano (2015): "Culture and Institutions," *Journal of Economic Literature*, 53(4), 898–944.

Algan, Y., and P. Cahuc (2010): "Inherited Trust and Growth," *American Economic Review*, 100, 2060–2092.

Alsan, M. (2015): "The Effect of the TseTse Fly on African Development," *American Economic Review*, 105(1), 382–410.

Antweiler, W., and M. Z. Frank (2004): "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *Journal of Finance*, 59(3), 1259–1294.

Ashraf, Q., and O. Galor (2017): "The Macrogenoeconomics of Comparative Development," Journal of Economic Literature, forthcoming.

Baker, S. R., N. Bloom, and S. J. Davis (2016): "Measuring Economic Policy Uncertainty," *Quarterly Journal of Economics*, 131(4), 1593–1636.

Bascom, W. R. (1953): "Folklore and Anthropology," *Journal of American Folklore*, 66(262), 283–290.

Bates, R. H. (1983): *Essays on the Political Economy of Rural Africa*. University of California Press, Berkeley, CA.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain," *Econometrica*, 80(6), 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on Treatment Effects After Selection Among High-Dimensional Controls," *Review of Economic Studies*, 81, 608–650.

——— (2015): "High-dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economic Perspectives*, 28(2), 608–650.

Benedict, R. (1934): *Patterns of Culture*. Houghton Mifflin Harcourt, New York.

Bentzen, J., J. G. Hariri, and J. A. Robinson (2015): "The Indigenous Roots of Representative Democracy," NBER Working Paper 21193.

Bentzen, J. S. (2015): "Acts of God: Religiosity and Natural Disasters Across Subnational World Districts," University of Copenhagen Working Paper.

Berelson, B. (1952): *Content Analysis in Communications Research*. The Free Press.

Berezkin, Y. (2015a): "Folklore and Mythology Catalogue: its Lay-out and Potential for Research," in *The Retrospective Methods Network Newsletter*, ed. by Frog, H. F. Leslie-Jacobsen, and J. S. Hopkins, vol. 10, chap. Between Text and Practice: Mythology, Religion and Research, pp. 58–70. Folklore Studies / Dept. of Philosophy, History, Culture and Art Studies University of Helsinki, Helsinki, Helsinki.

——— (2015b): "Spread of Folklore Motifs as a Proxy for Information Exchange: Contact Zones and Borderlines in Eurasia," *Trames: Journal of Humanities and Social Sciences*, 19(1), 3–13.

——— (2016): "Peopling of the New World in Light of the Data on Distribution of Folklore Motifs," in *Maths Meets Myths: Quantitative Approaches to Ancient Narratives (Understanding Complex Systems)*, ed. by R. Kenna, M. M. Carron, and P. M. Carron, vol. 10, pp. 71–89. Springer Verlag.

Berezkin, Y., and E. Duvakin (2016): "Buried in a Head: African and Asian Parallels to Aesop's Fable," *Folklore*, 127(1), 91–102.

Boas, F. (1916): "The Development of Folk-Tales and Myths," *The Scientific Monthly*, 3(4), 335–343.

——— (2002): in *Indian Myths and Legends from the North Pacific Coast of America: A Translation of Franz Boas' 1895 Edition of Indianische Sagen von der Nordpazifischen Kuste Amerikas, trans. Dietrich Bertz*, ed. by R. Bouchard, and D. KennedyTalonbooks, Vancouver, Canada.

Bockstette, V., A. Chanda, and L. Putterman (2002): "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, 7(4), 347–369.

Bollen, J., and H. Mao (2011): "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, 2(1), 1–8.

Bolton, C., R. Bolton, L. Gross, A. Koel, C. Michelson, R. L. Munroe, and R. H. Munroe (1976): "Pastoralism and Personality: An Andean Replication," *Ethos*, 4(4), 463–481.

Bénabou, R., A. Falk, and J. Tirole (2018): "Narratives, Imperatives, and Moral Reasoning," NBER Working Paper No. 24798.

Cambria, E., A. Hussain, C. Havasi, and C. Eckl (2010): "SenticSpace Visualizing Opinions and Sentiments in a Multi-Dimensional Vector Space," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pp. 385–393. Springler.

Campbell, J. (1968): *The Hero with a Thousand Faces*. Princeton University Press, Novato, California.

Cecil, D. J., D. E. Buechler, and R. J. Blakeslee (2014): "Gridded Lightning Climatology from TRMM-LIS and OTD: Dataset Description," *Atmospheric Research*, 135-136, 404–414.

Cervellati, M., G. Chiovelli, and E. Esposito (2017): "Bite and Divide: Ancestral Exposure to Malaria and the Emergence and Persistence of Ethnic Diversity in Africa," Working paper, London Business School.

Chen, K. (2013): "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets," *American Economic Review*, 103(2), 690–731.

Dell, M., N. Lane, and P. Querubin (2017): "The Historical State, Local Collective Action, and Economic Development in Vietnam," working paper, Harvard University, Department of Economics.

Diamond, J., and J. A. Robinson (2010): *Natural Experiments of History*. Harvard University Press, Cambridge, MA.

Dorson, R. M. (1982): *Folklore and folklife: An introduction*. University of Chicago Press.

Dorst, J. D. (2016): "Folklore's Cybernetic Imaginary, or, Unpacking the Obvious," *Journal of American Folklore*, 129(512), 127–145.

Dundes, A. (1997): "The Motif-Index and the Tale Type Index: A Critique," *Journal of Folklore Research*, 34(3), 195–202.

Edgerton, R. D. (1971): "The Individual in Cultural Adaptation: a Study of Four East African Peoples," in *Studies in Culture and Ecology in East Africa*, ed. by W. GoldSchmidt. University of California Press, Berkeley and London.

El-Shamy, H. M. (2004): *Types of the Folktale in the Arab World: A Demographically Oriented Tale-Type Index*. Indiana University Press, Bloomington.

Enke, B. (2017): "Moral Values and Trump Voting," working paper, Harvard University.

——— (2018): "Kinship, Cooperation, and the Evolution of Moral Systems," Quarterly Journal of Economics, forthcoming.

Evans, J. A., and P. Aceves (2016): "Machine Translation: Mining Text for Social Theory," *Annual Review of Sociology*, 42, 21–50.

Fenske, J. (2013): "Does Land Abundance Explain African Institutions?," *Economic Journal*, 123(573), 1363–1390.

——— (2014): "Ecology, Trade, and States in Pre-colonial Africa," *Journal of the European Economic Association*, 12(3), 612–640.

Fernandez, R. (2011): "Does Culture Matter?," in *Handbook of Social Economics*, ed. by J. Benhabib, M. O. Jackson, and A. Bisin, vol. 1A, pp. 481–510. North-Holland, Amsterdam.

Fine, G. A. (1979): "Folklore Diffusion Through Interactive Social Networks: Conduits in a Preadolescent Community," *New York Folklore Society*, 5(1-2).

Fisman, R., and E. Miguel (2007): "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets," *Journal of Political Economy*, 115(6), 1020–1048.

Fortes, M., and E. Evans-Pritchard (1940): *African Political Systems*. Oxford University Press, New York.

Foucault, M. (1995): *Discipline and Punish: The Birth of the Prison*. Vintage Books, New York.

Galor, O., and O. Ozak (2015): "Land Productivity and Economic Development: Caloric Suitability vs. Agricultural Suitability," mimeo, Brown University.

Galor, O., O. Ozak, and A. Sarid (2016): "Geographical Origins and Economic Consequences of Language Structures," Brown University Working Paper.

Gebner, G. (1969): "The Analysis of Communication Content; Developments in Scientific Theories and Computer Techniques," Discussion paper.

Gentzkow, M., B. T. Kelly, and M. Taddy (2017): "Text as Data," NBER Working Paper 23276.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2016): "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech," NBER Working Paper 22423.

Giuliano, P., and N. Nunn (2013): "The Transmission of Democracy: From the Village to the Nation State," *American Economic Review, Papers and Proceedings*, 103(3), 86–92.

——— (2017): "Understanding Cultural Persistence and Change," NBER Working Paper No. 23617.

Goldberg, C. (1984): "The Historic-Geographic Method: Past and Future," *Journal of Folklore Research*, 21(1), 1–18.

GoldSchmidt, W. (1971): "Introduction: The Theory of Cultural Adaptation, The Individual in Cultural Adaptation: A Study of Four East African Peoples," in *Studies in Culture and Ecology in East Africa*, ed. by W. GoldSchmidt. University of California Press, Berkeley and London.

Gottschalk, L. A., and G. C. Gleser (1969): *The Measurement of Psychological States through the Content Analysis of Verbal Behavior*. University of California Press.

Graham, J., J. Haidt, S. Koleva, M. Motyl, P. Iyer, S. Wojcik, and P. Ditto (2013): "Moral foundations theory: The pragmatic validity of moral pluralism," in *Advances in Experimental Social Psychology*, vol. 47, pp. 55–130. Elsevier.

Graham, J., J. Haidt, and B. A. Nosek (2009): "Liberals and conservatives rely on different sets of moral foundations.," *Journal of personality and social psychology*, 96(5), 1029.

Green, T. A. (1997): *Folklore: An Encyclopedia of Beliefs, Customs, Tales, Music, and Art*, vol. 1. ABC-CLIO, Santa Barbara, CA.

Grimmer, J., and B. Stewart (2013): "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis*, 21(3), 267–297.

Grosjean, P. (2014): "A History of Violence: The Culture of Honor and Homicide in the US South," *Journal of the European Economic Association*, 12(5), 1285–1316.

Haidt, J. (2007): "The New Synthesis in Moral Psychology," *Science*, 316(5827), 998–1002.

Haidt, J., and J. Craig (2004): "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues," *Daedalus*, 133(4), 55–66.

Harari, Y. N. (2015): *Sapiens: A Brief History of Humankind*. HarperCollins, New York.

Holsti, O. R. (1969): "Content Analysis for the Social Sciences and Humanities," .

Honti, J. (1939): "Marchenmorphologie und Marchen typologie," *Folk-Liv*, 3, 307–318.

Jason, H. (1970): "The Russian Criticism of the Finnish School in Folktale Scholarship," *Norveg*, 14, 285–294.

Jung, G. C. (1968): "The Archetypes and the Collective Unconscious," in *Collected Works of CG Jung*, vol. 9. Princeton University Press, Princeton, NJ.

Kaplan, D. (2000): "The Darker Side of the "Original Affluent Society"," *Journal of Anthropological Research*, 56(3), 301–324.

Landes, D. S. (1998): *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*. W.W. Norton, New York.

Levi-Strauss, C. (1955): "The Structural Study of Myth," *The Journal of American Folklore*, 68(270), 428–444.

Lowes, S., N. Nunn, J. A. Robinson, and J. L. Weigel (2017): "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom," *Econometrica*, 85(4), 1065–1091.

Mayshar, J., O. Moav, Z. Neeman, and L. Pascali (2015): "Cereals, Appropriability and Hierarchy," CEPR Discussion Paper 10742.

Michalopoulos, S. (2012): "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 102(4), 1508–1539.

Michalopoulos, S., A. Naghavi, and G. Prarolo (2018): "Trade and Geography in the Spread of Islam," *Economic Journal*, 128(616), 3210–3241.

Michalopoulos, S., and E. Papaioannou (2013): "Pre-colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 81(1), 113–152.

——— (2014): "National Institutions and Subnational Development in Africa," *Quarterly Journal of Economics*, 129(1), 151–213.

——— (2016): "The Long-Run Effects of the Scramble for Africa," *American Economic Review*, 106(7), 1802–1848.

——— (2017a): *The Long Economic and Political Shadow of History: Vol. 1: A Global View*chap. Introduction: Historical Legacies and Contemporary Development, pp. viii–xxix. VoxEU. PDF e-book.

——— (2017b): "Spatial Patterns of Development: A Meso Approach," Annual Review of Economics, forthcoming.

Michalopoulos, S., L. Putterman, and D. N. Weil (2018): "The Influence of Ancestral Lifeways on Individual Economic Outcomes in Sub-Saharan Africa," Journal of European Economic Association, forthcoming.

Murdock, G. P. (1959): *Africa: Its Peoples and their Culture History*. McGraw-Hill, New York.

——— (1967): *Ethnographic Atlas: A Summary*. University of Pittsburgh Press, Pittsburgh, PA.

Nisbett, R. E., and D. Cohen (1996): *Culture of Honor: The Psychology of Violence in the South*. Westview Press, Boulder, CO.

North, R. C. (1963): *Content Analysis: A Handbook with Applications for the Study of international Crisis*. Northwestern University Press.

Nunn, N. (2012): "Culture and the Historical Process," *Economic History of Developing Regions*, 27(S1), 108–126.

Nunn, N., and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101(7), 3221–3252.

Osafo-Kwaako, P., and J. A. Robinson (2013): "Political Centralization in Pre-colonial Africa," *Journal of Comparative Economics*, 41(1), 6–21.

Pennebaker, J. W., R. J. Booth, and M. E. Francis (2007): "Linguistic inquiry and word count: LIWC [Computer software]," *Austin, TX: liwc. net*.

Richerson, P. J., M. B. Mulder, and B. J. Vila (2001): *Principles of Human Ecology*. University of California, Davis, PDF e-book.

Ross, R. M., S. J. Greenhill, and Q. D. Atkinson (2013): "Population Structure and Cultural Geography of a Folktale in Europe," *Proceedings of the Royal Society B*, 280(1756).

Sahlins, M. (1972): *Stone Age Economics*. Aldine Publishing, Chicago.

Saiz, A., and U. Simonsohn (2013): "Proxying for Unobservable Variables with Internet Document-Frequency," *Journal of the European Economic Association*, 11(1), 137–165.

Shiller, R. J. (2017): "Narrative Economics," *American Economic Review*, 107(4), 967–1004.

Speer, R., and J. Lowry-Duda (2017): "Conceptnet at semeval-2017 task 2: Extending word embeddings with multilingual relational knowledge," *arXiv preprint arXiv:1704.03560*.

Spolaore, E., and R. Wacziarg (2009): "The Diffusion of Development," *Quarterly Journal of Economics*, 124(2), 469–529.

Spolaore, E., and R. Wacziarg (2013): "How Deep Are the Roots of Economic Development?," *Journal of Economic Literature*, 51(2), 325–369.

Tausczik, Y. R., and J. W. Pennebaker (2010): "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of language and social psychology*, 29(1), 24–54.

Tetlock, P. C. (2007): "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance*, 62(3), 1139–1168.

Thompson, S. (1946): *The Folktale*. The Dryden Press, New York, NY.

——— (1955): *Motif-Index of Folk-Literature: a Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Medieval Romances, Exempla, Fabliaux, Jest-Books and Local Legends*. Indiana University Press.

Toelken, B. (1996): *Dynamics of Folklore*. University Press of Colorado, Boulder, CO.

Ullberg, J., S. Coradeschi, and F. Pecora (2010): "On-line ADL Recognition with Prior Knowledge," in *Proceedings of the 2010 Conference on STAIRS 2010: Proceedings of the Fifth Starting AI Researchers' Symposium*, pp. 354–366, Amsterdam, The Netherlands, The Netherlands. IOS Press.

Uther, H.-J. (2004): *The Types of International Folktales: a Classification and Bibliography, Based on the System of Antti Aarne and Stith Thompson*. Suomalainen Tiedeakatemia, Helsinki, Finland.

Voigtlander, N., and H.-J. Voth (2012): "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany," *Quarterly Journal of Economics*, 127(1), 1339–1392.

Weber, E. (1976): *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914*. Stanford University Press, Redwood City, CA.

Wisniewski, T. P., and B. Lambe (2013): "The Role of Media in the Credit Crunch: The Case of the Banking Sector," *Journal of Economic Behavior and Organization*, 85(1), 163–175.

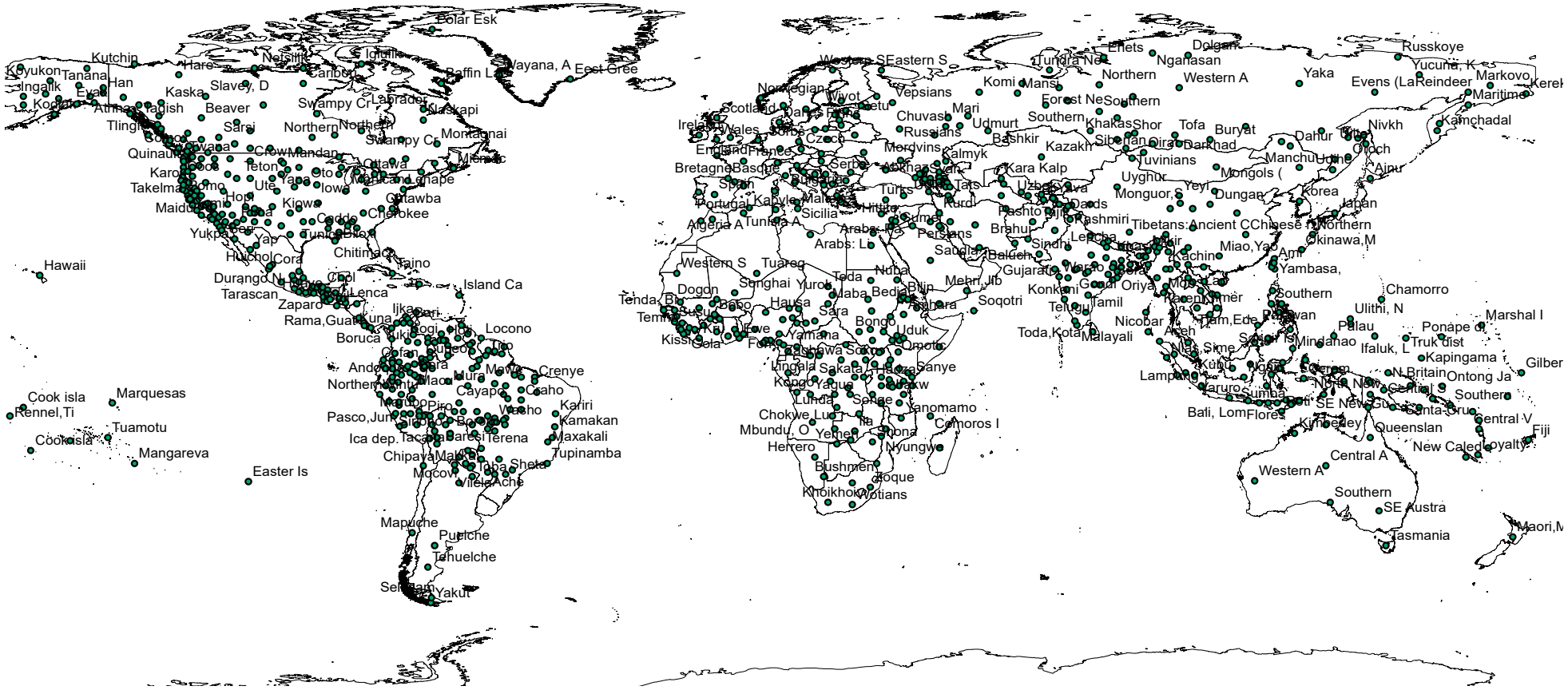
6 Appendix

6.1 The Construction of a Motif/Oral Tradition

Motif $k27n1$ is titled "Task-giver is a king or a chief" and the description reads: "Person who gives difficult tasks to the hero is a prominent figure in social hierarchy, i.e., a head of political unit of community - or higher level and not a mythical being." This motif is among the top-10 most widespread motifs being present in 228 societies (Appendix Figure 2 depicts its spatial distribution). Going briefly over the sources consulted by Berezkin to map the presence of this motif across oral traditions it becomes clear that each appearance of k27n1 is associated with a different tale/myth. Depending on the context, the "prominent figure" in the tale may be a king, queen, sultan, emir, tsar, rajah, emperor, boyar, khan, prince, aldar, padishah, official, mayor, chief, chieftain, or kurak. This implies that the use of "king" as a seed for capturing the presence of hierarchy across motifs takes into account many of the alternative configurations of political authority worldwide and not only those strictly related to the notion of kingdoms.

For the Telugu (15th largest language group worldwide), for example, Berezkin consulted a range sources to construct its oral tradition. Those sources include 9 different authors and were written in 3 languages which is typical of the societies in his catalogue. Altogether, 53 motifs can be found in the Telugu oral tradition. The particular tale where motif $k27n1$ can be found comes from the book "Folktales from India" written by Mosur Venkataswami and published by the Diocesan Press publishing house in Madras in 1923. Specifically, it is in the tale: "The adventures of Ratnalpolchetty," where a prince gets to perform a series of difficult tasks, including finding the Parijata flower and rare gems upon request of the ruler's daughter in order to convince her that he is worthy of her. This type of story is very much in the spirit of what Campbell in his famous 1968 work "The Hero with a Thousand Faces" described as the hero's journey, an archetypical narrative that involves a hero who goes on an adventure, wins decisive victories, and returns home transformed.

Other works by Mosur Venkataswami include "Five hundred instructive Tamil proverbs with their English equivalents" and "Uttama neeti". Panels A and B in Appendix Table 1 describe the underlying sources consulted by Berezkin for the Telugu oral tradition and the resulting motifs, respectively.


Polar Esk
Kutchin
Koyukon
Tanana
Ingalik
Eyak
Han
Kaska
Slavey, D
Hare
Kodiak
Athna
Tadish
Beaver
Tlingit
Sarsi
Swampy Cr
Labrador
Slaskapi
Northern
Montagnai
Quinault
Crow
Mandan
Teton
Yana
Oto
Iowa
Ottawa
Chippewa
Menomini
Ojibwa
Cherokee
Karok
Takelma
Ute
Hopi
Kiowa
Maidu
Yokuts
Yuki
Cahto
Yuki
Pomo
Yuma
Cocopa
Mohave
Diegueno
Luiseno
Cupeno
Cahuilla
Serrano
Gabrielino
Chumash
Tubatulabal
Kawaiisu
Kitanemuk
Tataviam
Vanyume
Desert Cahuilla
Cocopa
Halchidhoma
Kohuana
Maricopa
Mohave
Yavapai
Walapai
Havasupai
Hualapai
Coconino
Southern Paiute
Chemehuevi
Kaibab
Uintah Ute
Gosiute
Shoshoni
Bannock
Northern Paiute
Washo
Northern Paiute
Klamath
Modoc
Shasta
Achomawi
Atsugewi
Wintun
Wintu
Yana
Maidu
Nomlaki
Patwin
Lake Miwok
Yuki
Pomo
Coast Miwok
Esselen
Salinan
Chumash
Yokuts
Tubatulabal
Kawaiisu
Kitanemuk
Tataviam
Serrano
Gabrielino
Luiseno
Diegueno
Cupeno
Cahuilla
Cocopa
Halchidhoma
Kohuana
Maricopa
Mohave
Yavapai
Walapai
Havasupai
Hualapai
Coconino
Southern Paiute
Chemehuevi
Kaibab
Uintah Ute
Gosiute
Shoshoni
Bannock
Northern Paiute
Washo

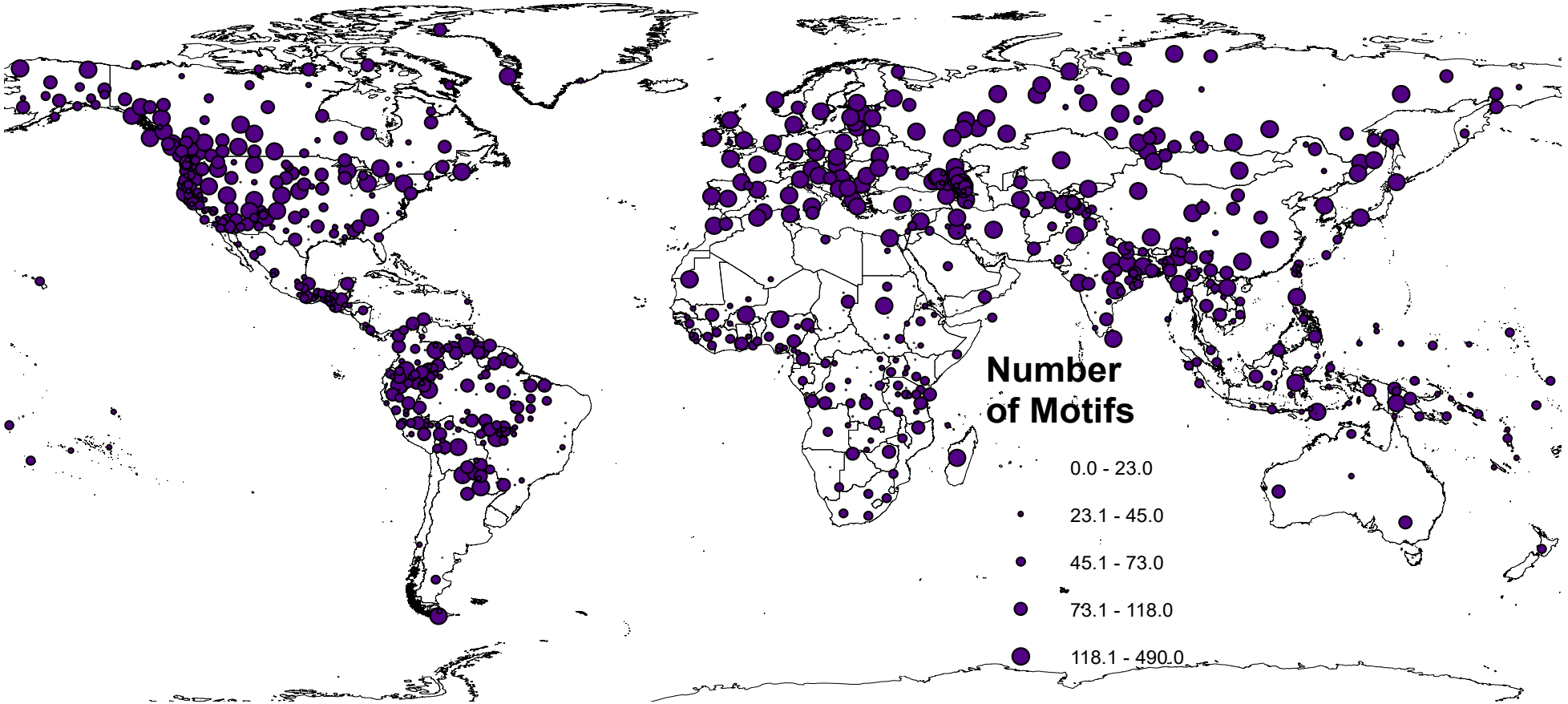


Figure 2: # of Motifs across Oral Traditions

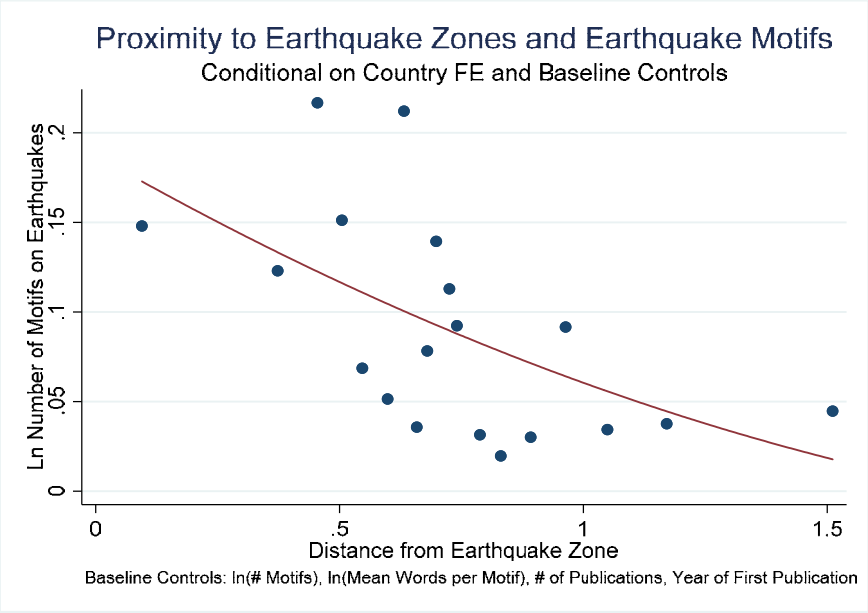


Figure 3a

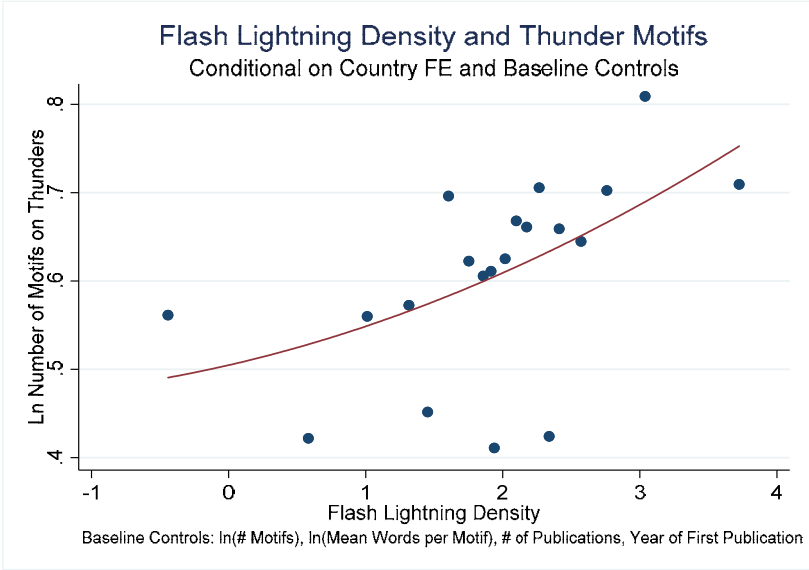


Figure 3b

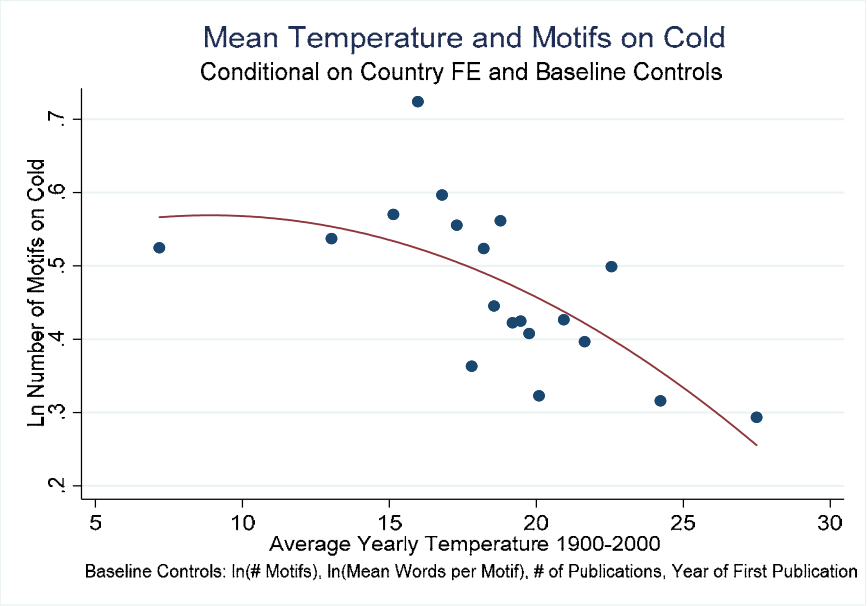


Figure 3c

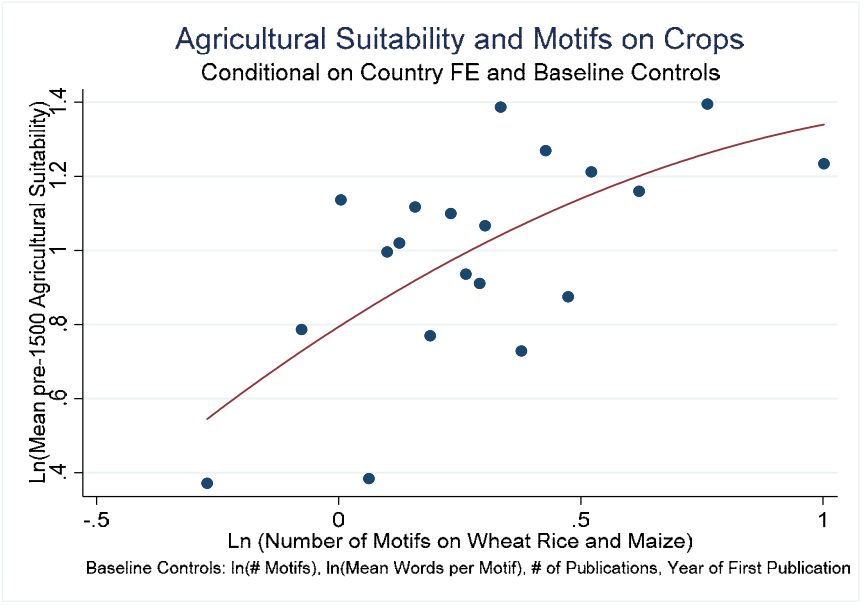


Figure 3d

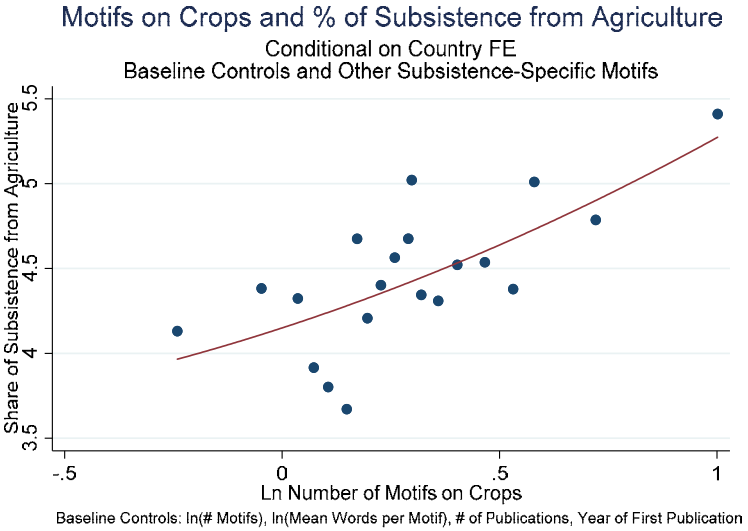


Figure 4a

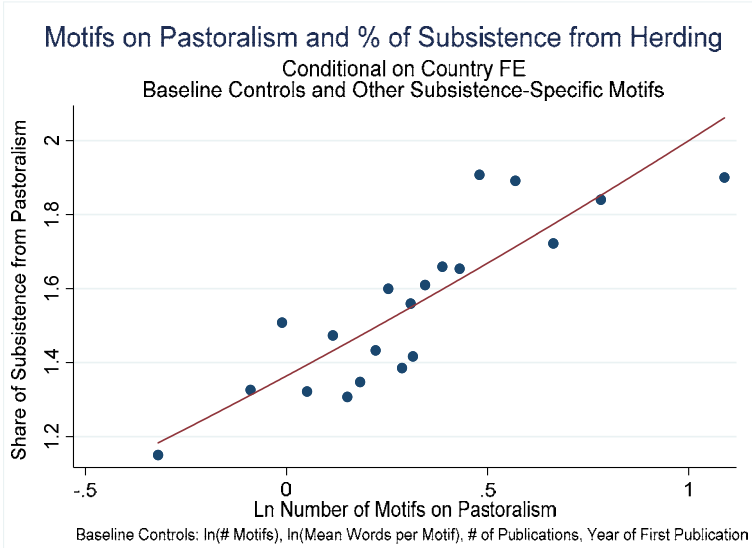


Figure 4b

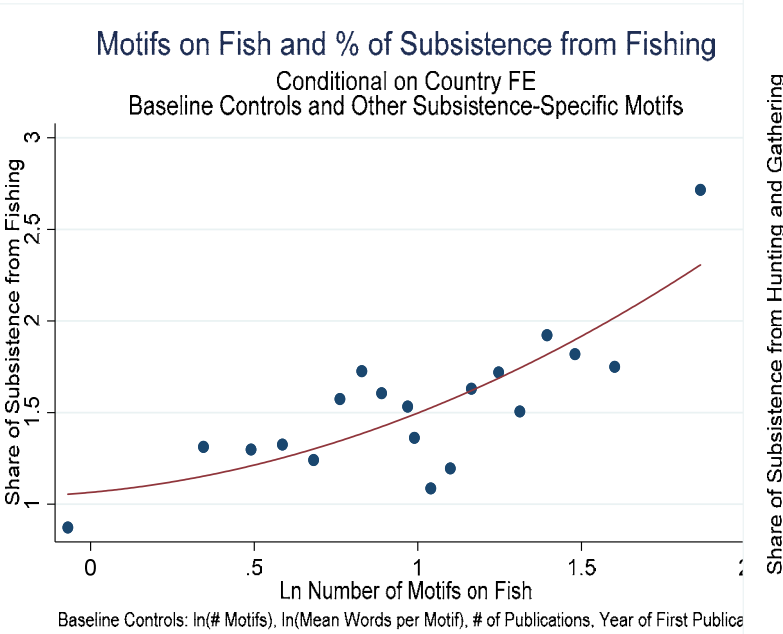


Figure 4c

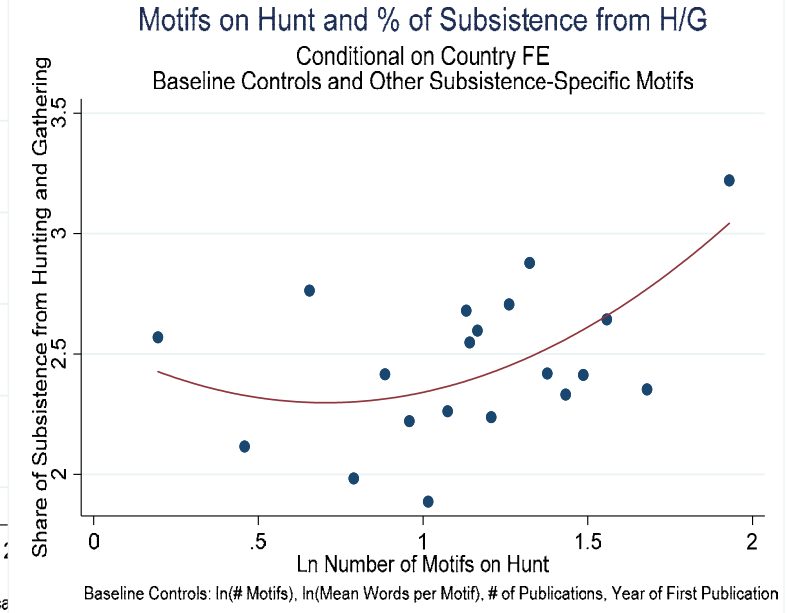


Figure 4d

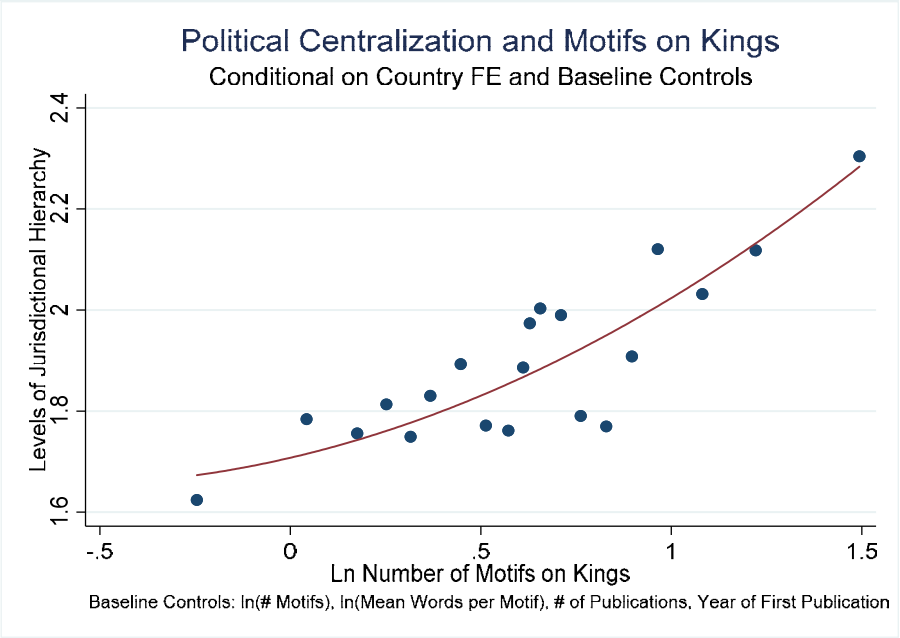


Figure 5a

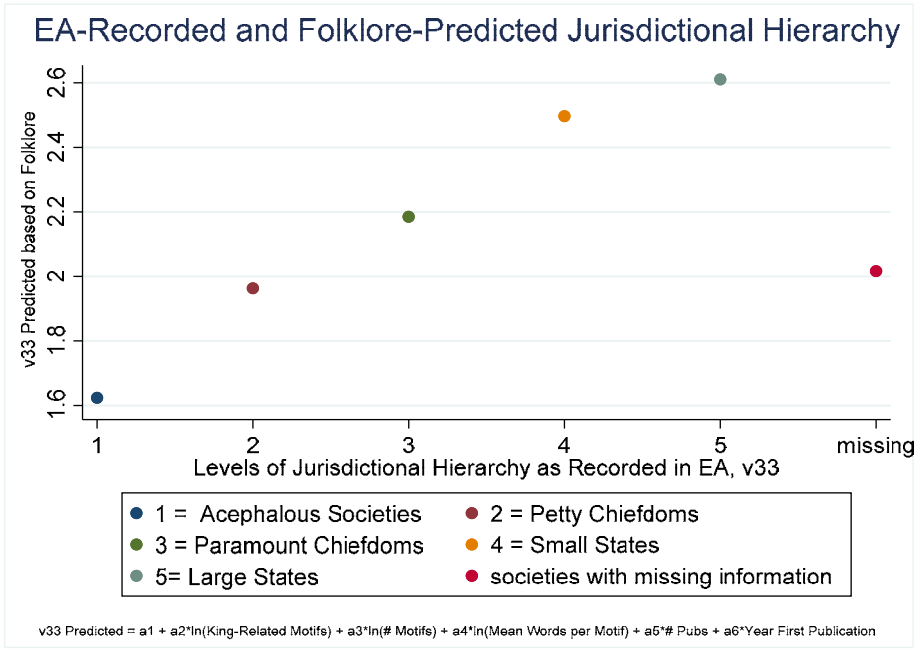


Figure 5b

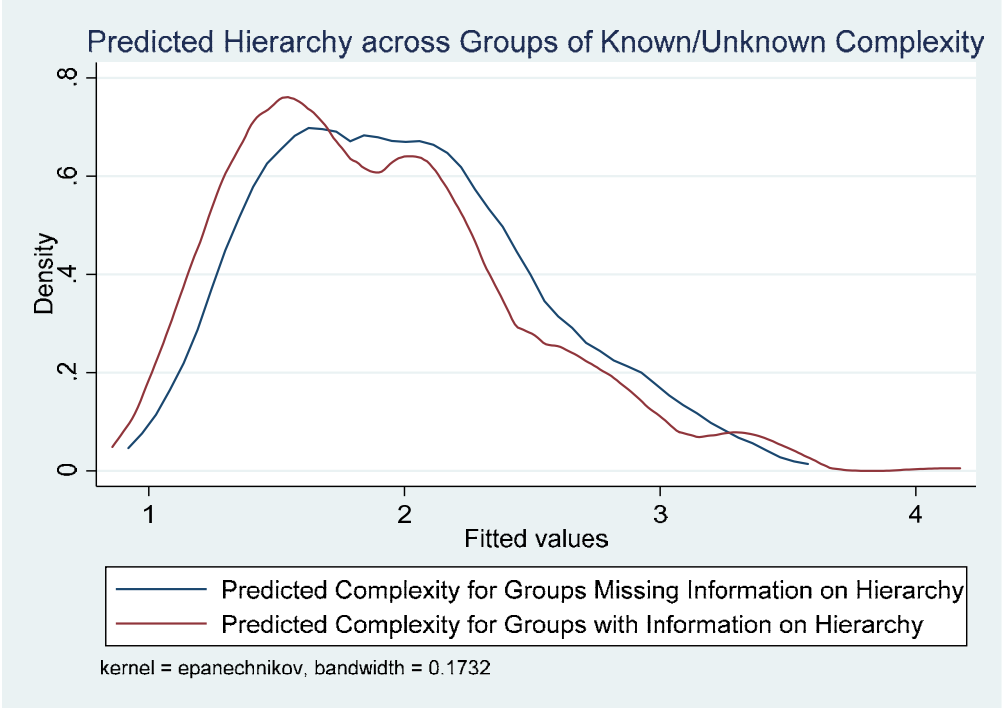


Figure 5c

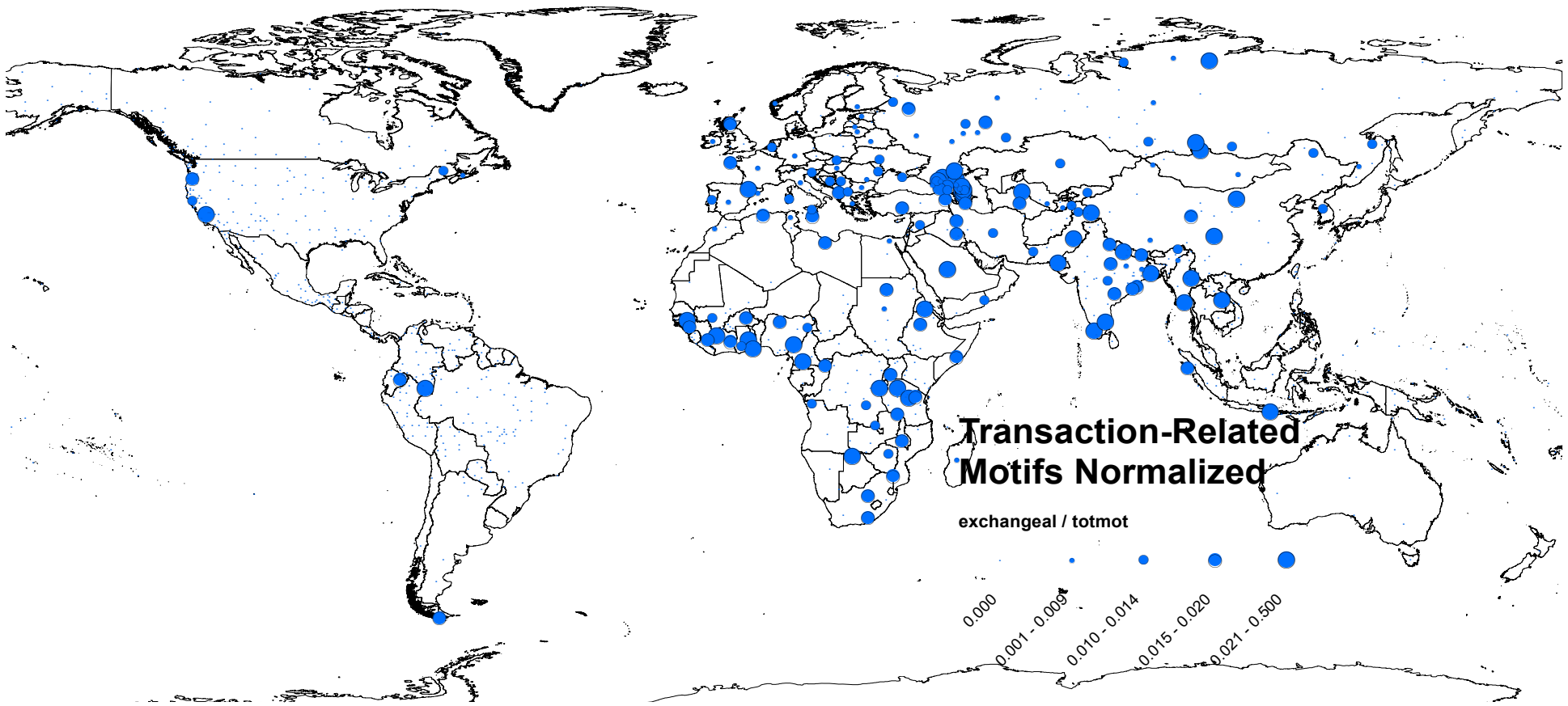


Figure 6: # of Trade-Related Motifs across Oral Traditions

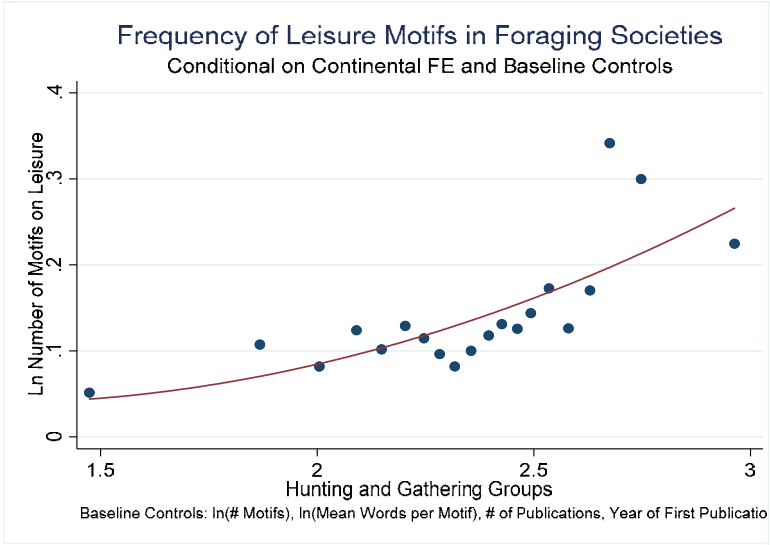


Figure 7a

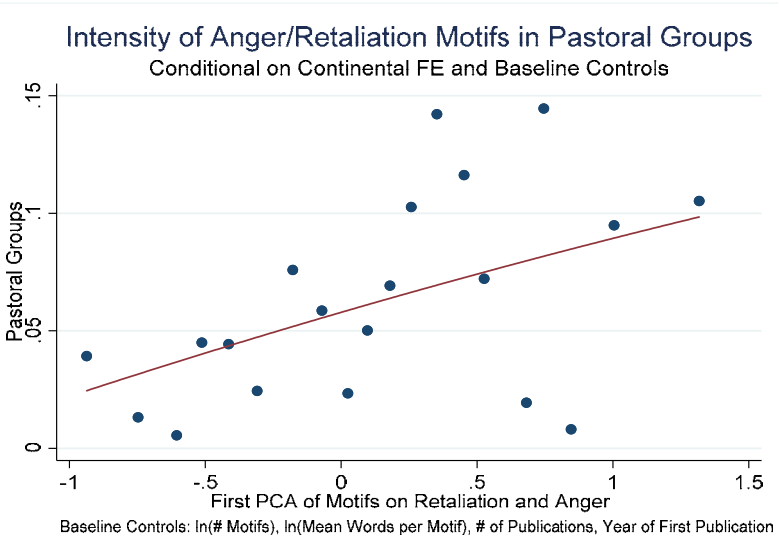


Figure 7b

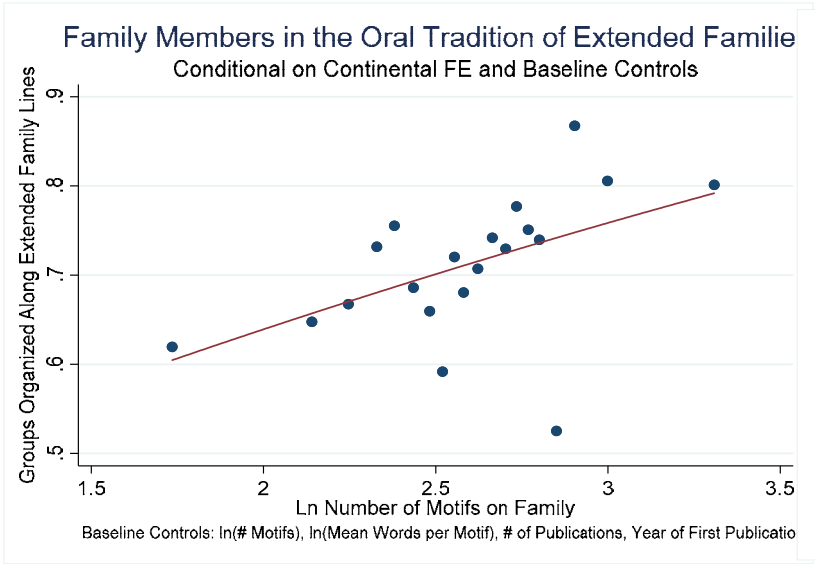


Figure 7c

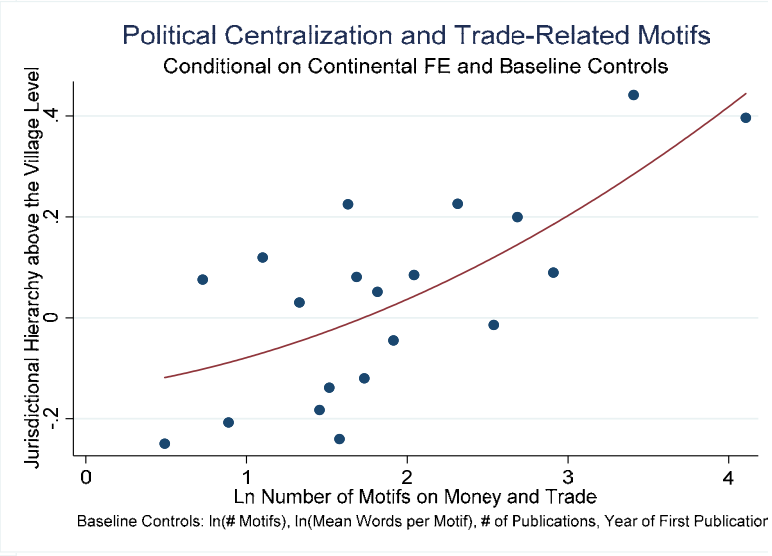


Figure 7d

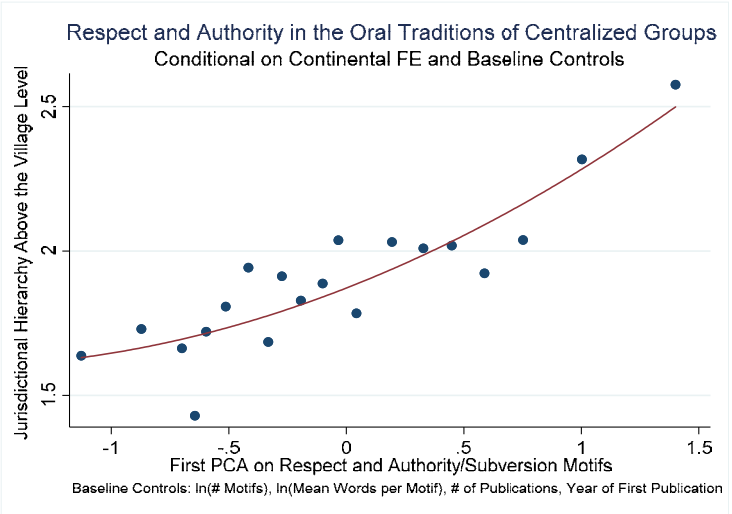
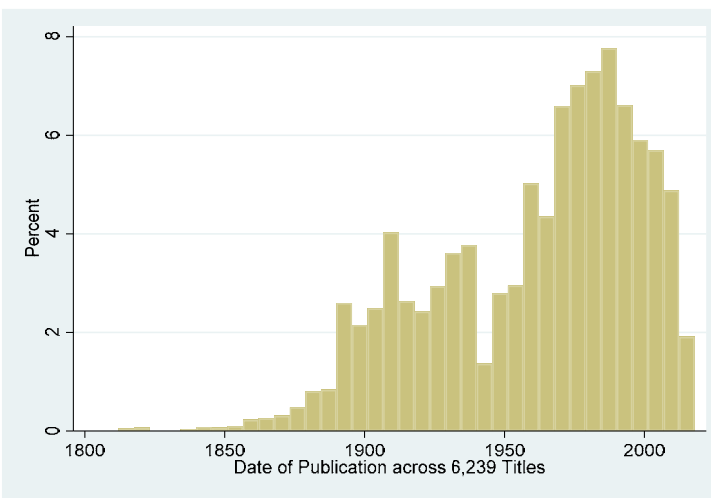
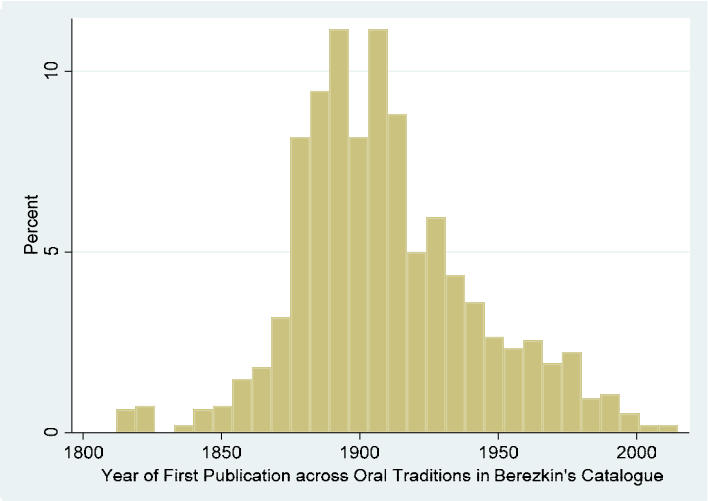


Figure 7e



Appendix Figure 1a
Date of Publication across 6,239 Titles



Appendix Figure 1b
Date of First Publication across Groups

Appendix Figure 2: The Spatial Distribution of Motif k27n1

Motif Description: "Person who gives difficult tasks to the hero is a prominent figure in social hierarchy, i.e., a head of political unit of community - or higher level and not a mythical being"

Table 1 - Panel A: Language Breakdown of Publications in Berezkin's Catalogue

Language	# of Publications	Language	# of Publications
English	2527	Belarusian	7
Russian	1844	Scots Gaelic	6
Spanish	558	Czech	4
German	452	Arabic	2
French	389	Galician	2
Portuguese	76	Luxembourgish	2
Estonian	49	Serbian	2
Romanian	42	Ukrainian	2
Hungarian	40	Bulgarian	1
Danish	39	Samoan	1
Dutch	30	Shona	1
Nepali	28	Welsh	1
Catalan	22		
Polish	21		
Latin	18		
Italian	16		
Slovenian	16		
Swedish	16		
Bosnian	13		
Finnish	10		

Table 1 - Panel B: Top 10 Groups with the Largest Number of Motifs

Group	# of Motifs
Finns	417
Greek (modern)	429
Germans	429
Latvians	437
Kazakh	439
Lithuanians	457
Georgians	465
Bulgaria	533
Ukrainians	556
Russians	563

Table 1 - Panel C: Examples of Motifs in Each of the 12 categories of Berezkin's Catalogue

Motif Group	ID	Description
a	a3	The Moon is female or bisexual, the Sun is male.
b	a32	A figure or an imprint of some being or object are seen in the Moon.
c	b3a	Water is the original element, the dry earth appears later.
d	h28	Killed and destroyed (often burned) person or creature (usually ogre, fierce animal, powerful shaman) turns into a multitude of biting insects or into other small molesting creatures.
e	f9	For different reasons, sexual contact with a woman is deadly dangerous for a man.
f	g6	One of the trees is the principal, original one (emerged before all the other; ancestor of wild or cultivated plants; ocean or rivers inside it; world axis; higher than all the others; overshadows sky).
g	b82	Raven or other carrion-eating bird of dark color and a similar size was originally white.
h	l19b	A being with three or more heads is described in tales or represented in art.
i	m29b	In episodes related to deception, absurd, obscene or anti-social behavior the protagonist is fox, jackal or coyote.
j	k27n	Father or other kinsmen of hero's wife or bride try to kill or test him and/or suggest him difficult tasks.
k	m30	Person or creature who has no wings or is unable to fly on a long distance attempts to ascend to the sky or to fly far away but falls down or, deprived of his wings, remains in a place from which he is unable to return.
l	k38e	Loci or objects of three (rare – four) different materials are mentioned in such a way that all of them have positive connotations though unequal value (copper, silver and gold; silver, gold and diamonds, etc.)

Table 2 - Panel A: Summary Statistics for the Oral Traditions in Berezkin's Catalogue

	mean	p50	sd	min	max	N
# of Motifs	84.05	59.00	83.18	1.00	563.00	945
Word Count of an Average Motif	60.03	58.95	7.49	33.00	89.67	945
# of Publications	14.84	10.00	13.40	1.00	104.00	945
# of Languages in the Publications	3.53	3.00	2.44	1.00	14.00	945
# of Authors	12.48	9.00	11.30	1.00	84.00	945
Year of First Publication	1908.19	1904.00	35.61	1638.00	2015.00	945
Year of Average Publication	1958.65	1960.19	19.56	1882.00	2015.00	945
# of Motifs on Earthquakes	0.10	0.00	0.34	0.00	3.00	945
# of Motifs on Thunder	1.21	1.00	1.50	0.00	9.00	945
# of Motifs on Cold	0.85	0.00	1.24	0.00	6.00	945
# of Motifs on Wheat, Rice and Maize	0.58	0.00	1.07	0.00	8.33	945
Distance to Earthquake Zones of Strength 3 and 4	575.94	315.41	716.39	0.00	4694.88	945
Mean Lightning Flash Rates (1995-2010)	10.61	6.72	11.48	0.05	96.87	925
Mean Annual Temperature (1901-2000)	14.78	17.77	11.40	-20.31	30.41	891
Optimal Agricultural Calories Pre-1500	5960.03	6345.41	4040.19	1.00	19175.24	942
Change in Optimal Agricultural Calories from the Columbian Exchange	1369.04	503.28	1883.07	-230.00	10519.32	942

See Main Text for variable definition

Table 2 - Panel B: Summary Statistics of Oral Traditions and Ethnographic Traits (in the Sample of the EA)

	mean	p50	sd	min	max	N
# of Motifs	81.03	63.00	69.48	2.00	563.00	1,238
Word Count of an Average Motif	60.75	59.70	6.88	38.91	80.77	1,238
# of Publications	16.34	14.00	11.54	1.00	104.00	1,238
# of Languages in the Publications	3.80	4.00	2.17	1.00	14.00	1,238
# of Authors	13.80	12.00	9.71	1.00	84.00	1,238
Year of First Publication	1906.12	1905.00	28.52	1638.00	2015.00	1,238
Year of Average Publication	1956.85	1956.56	15.96	1882.00	2015.00	1,238
% of Subsistence from Hunting and Gathering	2.46	1.00	2.73	0.00	10.00	1,264
Predominantly Hunting and Gathering Groups	0.14	0.00	0.35	0.00	1.00	1,264
% of Subsistence from Fishing	1.53	1.00	1.70	0.00	9.00	1,264
% of Subsistence from Animal Husbandry	1.56	1.00	1.80	0.00	9.00	1,264
Predominantly Pastoral Groups	0.06	0.00	0.24	0.00	1.00	1,264
% of Subsistence from Agriculture	4.45	5.00	2.71	0.00	9.00	1,264
Levels of Jurisdictional Hierachy Beyond the Village level	1.91	2.00	1.05	1.00	5.00	1,129
# of Motifs on Trade (ConceptNet)	1.24	0.00	2.05	0.00	14.00	1,238
# of Motifs on Money (LIWC)	6.61	3.00	10.22	0.00	77.00	1,238
First PCA of Motifs on Trade and Money	0.06	-0.22	1.24	-1.50	4.15	1,238
# of Motifs on Wheat, Maize, and Rice (ConceptNet)	0.50	0.00	0.90	0.00	8.33	1,238
# of Motifs on Pastoralism (ConceptNet)	0.72	0.00	1.77	0.00	15.00	1,238
# of Motifs on Fish (ConceptNet)	2.53	2.00	2.94	0.00	19.00	1,238
# of Motifs on Hunt (ConceptNet)	3.19	2.00	3.57	0.00	29.00	1,238
# of Motifs on King (ConceptNet)	1.81	1.00	3.70	0.00	23.00	1,238
Distance to Pre-600AD Trade Routes in 1000km	3.14	2.04	2.69	0.00	11.75	1,238
Change in Distance to Trade Routes Between 600 AD and 1700AD	-0.33	-0.06	1.24	-2.87	1.61	1,238
Group Organized Along Extended Family Lines	0.71	1.00	0.45	0.00	1.00	1,237
# of Motifs on Leisure (LIWC)	13.71	10.00	12.83	0.00	93.00	1,238
# of Motifs on Family (LIWC)	19.18	13.00	21.25	0.00	158.00	1,238
# of Motifs on Anger (LIWC)	27.78	20.00	27.41	0.00	210.00	1,238
# of Motifs on Retaliation (ConceptNet)	1.21	1.00	1.47	0.00	9.00	1,238
First PCA of Motifs on Anger and Retaliation	0.13	0.11	1.22	-2.52	3.64	1,238
# of Motifs on Respect (ConceptNet)	0.74	0.00	1.46	0.00	10.00	1,238
# of Motifs on Respect to Authority (MFD)	9.29	6.00	11.05	0.00	79.00	1,238
First PCA of Motifs on Respect and Respect for Authority	-0.05	-0.28	1.12	-1.78	3.93	1,238

See Main Text for variable definition. In parentheses we mention the dictionary used to tag the motifs. These are the: ConceptNet, LIWC and MFD, respectively.

Table 3: Folklore, Subsistence, and the Physical Environment

	Ln(1+# Motifs on Earthquakes)		Ln(1+# Motifs on Thunder)		Ln(1+# Motifs on Cold)		Ln(1+# Motifs on Crops)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance to Earthquake Zones in 1000 km's	-0.0524***	-0.0650***							
	(0.0097)	(0.0165)							
Ln(Mean Lightning Flash Density)			0.0337**	0.0424***					
			(0.0156)	(0.0159)					
Mean Yearly Temperature 1900-2000					-0.0108***	-0.0070*			
					(0.0022)	(0.0042)			
Ln(Optimal Agricultural Calories Pre-1500)							0.0196***	0.0177**	0.0158**
							(0.0053)	(0.0067)	(0.0070)
Change in Optimal Agricultural Calories from the Columbian Exchange									0.0133
									(0.0088)
Baseline Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Continental FE	yes	no	yes	no	yes	no	yes	no	no
Country FE	no	yes	no	yes	no	yes	no	yes	yes
R^2	0.046	0.079	0.426	0.452	0.362	0.393	0.444	0.47	0.471
# of Observations	945	945	925	925	891	891	942	942	942

Notes: This table reports OLS estimates. The unit of analysis is a group in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # Publications, Year of First Publication. Standard errors are clustered at the language family level as recorded by Berezkin. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main text for variable definition and Table 2 - Panel A for summary statistics.

Table 4: Folklore, Subsistence, and the Ethnographic Record

Share of Subsistence from:	Farming		Animal Husbadry		Fishing		Hunting & Gathering	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(Farming Motifs + 1)	1.5333***	1.0540***	0.3067***	0.0673	-0.5912***	-0.4478***	-1.2340***	-0.6668***
	(0.2717)	(0.1734)	(0.0955)	(0.1475)	(0.1533)	(0.1381)	(0.2337)	(0.2117)
Ln(Pastoralism Motifs + 1)	-0.146	-0.2125	0.7669***	0.6241***	-0.1745*	-0.0269	-0.4656**	-0.4001***
	(0.1606)	(0.1921)	(0.1659)	(0.0925)	(0.0899)	(0.1479)	(0.1782)	(0.1099)
Ln(Fish-Related Motifs + 1)	-0.1505	-0.2211	-0.2227***	-0.1391	0.6414***	0.6550***	-0.2744	-0.3031
	(0.1823)	(0.2151)	(0.0739)	(0.0900)	(0.1499)	(0.1473)	(0.2571)	(0.2536)
Ln(Hunt-Related Motifs + 1)	-1.1383***	-0.3950**	0.4222**	0.2702**	-0.0645	-0.2141**	0.7802***	0.3452**
	(0.2445)	(0.1689)	(0.2071)	(0.1326)	(0.1273)	(0.1059)	(0.2155)	(0.1419)
Baseline Controls	yes	yes	yes	yes	yes	yes	yes	yes
Continental FE	yes	no	yes	no	yes	no	yes	no
Country FE	no	yes	no	yes	no	yes	no	yes
R^2	0.414	0.55	0.443	0.59	0.286	0.37	0.523	0.623
# of Observations	1237	1237	1237	1237	1237	1237	1237	1237

Notes: This table reports OLS estimates. The unit of analysis is a group in the Ethnographic Atlas which has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition and Table 2 - Panel B for summary statistics.

Table 5: Folklore, Institutions and Exchange

	Levels of Jurisdictional Hierarchy Beyond the Village		Ln(1+# Motifs on Trade)		
	(1)	(2)	(3)	(4)	(5)
Ln(King-Related Motifs+1)	0.4565***	0.3520***			
	(0.1151)	(0.0946)			
Distance to Pre-600AD Trade Routes			-0.1637***	-0.2436***	-0.2663***
			(0.0290)	(0.0542)	(0.0556)
Change in Distance to Trade Routes Between 600 AD and 1700AD					-0.1134
					(0.0996)
Baseline Controls	yes	yes	yes	yes	yes
Continental FE	yes	no	yes	no	no
Country FE	no	yes	no	yes	yes
R^2	0.316	0.424	0.6	0.596	0.597
# of Observations	1108	1108	474	474	474

Notes: This table reports OLS estimates. In Columns 1 and 2 the unit of analysis is a group in the Ethnographic Atlas which has been matched to an oral tradition in Berezkin's catalogue. In Columns 3, 4, and 5 the unit is a group in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Columns 3-5 focus on societies in the Old World. Standard errors in columns 1 and 2 are clustered at the language family level as classified in the EA, v98, whereas standard errors in columns 3, 4 and 5 are clustered at the language family as classified by Berezkin. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition and Table 2 - Panel B for summary statistics.

Table 6: Motifs on Leisure in Hunter-Gatherer Societies

	ln(1+ Motifs on Leisure)			
	(1)	(2)	(3)	(4)
Predominantly Hunting and Gathering Groups	0.1799***	0.1709***	0.1042***	0.1179***
	(0.0538)	(0.0480)	(0.0386)	(0.0338)
Baseline Controls	yes	yes	yes	yes
Additional Controls	no	yes	no	yes
Continental FE	yes	yes	no	no
Country FE	no	no	yes	yes
# of Observations	1237	1237	1237	1237

Notes: This table reports OLS estimates. The additional controls in columns 2, and 4 are selected by the "post-double selection" methodology of Belloni et al (2015) from the set of 52 BoW in the LIWC and MFD dictionaries (naturally the BoW on leisure is not considered). In the "post-double selection" specifications the baseline controls as well as the respective FE are partialled out first. The unit of analysis is a group in the Ethnographic Atlas. A group in the EA has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition and Table 2 - Panel B for summary statistics.

Table 7 - Panel A: Pastoralism and Motifs on Violence

	First Principal Component of Anger and Retaliation			
	(1)	(2)	(3)	(4)
Predominantly Pastoral Groups	0.2188***	0.0887**	0.1903***	0.1388***
	(0.0420)	(0.0347)	(0.0700)	(0.0427)
Baseline Controls	yes	yes	yes	yes
Additional Controls	no	yes	no	yes
Continental FE	yes	yes	no	no
Country FE	no	no	yes	yes
# of Observations	1237	1237	1237	1237

Notes: This table reports OLS estimates. The additional controls in columns 2, and 4 are selected by the "post-double selection" methodology of Belloni et al (2015) from the set of 52 BoW in the LIWC and MFD dictionaries (the BoW on anger is naturally not considered). In the "post-double selection" specifications the baseline controls as well as the respective FE are partialled out first. The unit of analysis is a group in the Ethnographic Atlas. A group in the EA has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition and Table 2 - Panel B for summary statistics.

Table 7 - Panel B: Extended Families and Motifs on Family Members

	Ln(1+Motifs on Family)			
	(1)	(2)	(3)	(4)
Groups Organized along Extended Family Lines	0.0740***	0.0360**	0.0813***	0.0432**
	(0.0224)	(0.0180)	(0.0287)	(0.0194)
Baseline Controls	yes	yes	yes	yes
Additional Controls	no	yes	no	yes
Continental FE	yes	yes	no	no
Country FE	no	no	yes	yes
# of Observations	1211	1211	1211	1211

Notes: This table reports OLS estimates. The additional controls in columns 2, and 4 are selected by the "post-double selection" methodology of Belloni et al (2015) from the set of 52 BoW in the LIWC and MFD dictionaries (the BoW on family is naturally not considered). In the "post-double selection" specifications the baseline controls as well as the respective FE are partialled out first. The unit of analysis is a group in the Ethnographic Atlas. A group in the EA has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition and Table 2 - Panel B for summary statistics.

Table 8 - Panel A: States and the Exchange-Economy

	First Principal Component of Trade and Money					
	(1)	(2)	(3)	(4)	(5)	(6)
Levels of Jurisdictional Hierarhy	0.1532***	0.1306***	0.1175***	0.1446***	0.1388***	0.1317***
	(0.0413)	(0.0364)	(0.0355)	(0.0439)	(0.0388)	(0.0382)
Share of Subsistence from Animal Husbandry			0.0563***			0.0580***
			(0.0133)			(0.0140)
Share of Subsistence from Agriculture			-0.0003			-0.0055
			(0.0107)			(0.0126)
Baseline Controls	yes	yes	yes	yes	yes	yes
Additional Controls	no	yes	yes	no	yes	yes
Continental FE	yes	yes	yes	no	no	no
Country FE	no	no	no	yes	yes	yes
# of Observations	1108	1108	1108	1108	1108	1108

Notes: This table reports OLS estimates. The additional controls in columns 2, 3 and 5, 6 are selected by the "post-double selection" methodology of Belloni et al (2015) from the set of 52 BoW in the LIWC and MFD dictionaries (the BoW on money is naturally not considered). In the "post-double selection" specifications the baseline controls as well as the respective FE are partialled out first. The unit of analysis is a group in the Ethnographic Atlas. A group in the EA has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition and Table 2 - Panel B for summary statistics.

Table 8 - Panel B: Centralized Societies and Motifs on Respect for Authority

	First Principal Component of Authority and Respect					
	(1)	(2)	(3)	(4)	(5)	(6)
					High Gods	
					Absent	*Present*
Levels of Jurisdictional Hierarhy	0.1737***	0.0827***	0.1183***	0.0854***	0.0694**	0.1206***
	(0.0430)	(0.0277)	(0.0316)	(0.0204)	(0.0302)	(0.0278)
Baseline Controls	yes	yes	yes	yes	yes	yes
Additional Controls	no	yes	no	yes	yes	yes
Continental FE	yes	yes	no	no	no	no
Country FE	no	no	yes	yes	yes	yes
# of Observations	1108	1108	1108	1108	556	158

Notes: This table reports OLS estimates. The controls in columns 2, 4, 5, 6 are selected by the "post-double selection" methodology of Belloni et al (2015). The additional controls set includes the 52 BoW from the LIWC and MFD dictionaries (the anger-related BoW from LIWC is naturally not included). The unit of analysis is a group in the Ethnographic Atlas. A group in the EA has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. In the "post-double selection" specifications in columns 2, 4, 5, 6 the baseline controls are partialled out. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition and Table 2 - Panel B for summary statistics.

Table 9: Oral Traditions and Current Attitudes

Panel A: Across Groups

Group Mean to the Question:	A Respondent finds Justifiable: Claiming Govt Benefits (F114)	Avoiding a Fare on Public Transport (F115)	Cheating on Taxes (F116)	Someone Accepting a Bribe (F117)	How much you trust: Your family (D001_B)		It is Important to this Person to Have a Good Time (A192)	People Who Don't Work Turn Lazy (C038)	Justifiable: Violence Against Other People (F114_03)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First PCA of Respect and Respect for Authority BoWs	-0.4040***	-0.3353**	-0.3245**	-0.4991***					
	(0.1389)	(0.1475)	(0.1457)	(0.1318)					
Ln(# of Motifs on Family)					-0.1327**	-0.1243**			
					(0.0593)	(0.0597)			
Ln(# of Motifs on Leisure)							-0.3944*	0.2770*	
							(0.2157)	(0.1644)	
First PCA of Retaliation and Anger BoWs									0.0586**
									(0.0234)
Baseline Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.119	0.093	0.035	0.152	0.105	0.103	0.111	0.027	0.143
# of Observations	163	164	161	165	114	114	118	116	84

Panel B: Across Individuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First PCA of Respect and Respect for Authority BoWs	-0.5274***	-0.3794**	-0.3628**	-0.4542***					
	(0.1536)	(0.1478)	(0.1478)	(0.1468)					
Ln(# of Motifs on Family)					-0.1394**	-0.1512**			
					(0.0593)	(0.0625)			
Ln(# of Motifs on Leisure)							-0.3844***	0.2148**	
							(0.1184)	(0.0937)	
First PCA of Retaliation and Anger BoWs									0.0259*
									(0.0146)
Baseline Controls, Country FE, Individual-Level Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.094	0.104	0.085	0.09	0.056	0.057	0.172	0.093	0.105
# of Observations	296094	295484	295542	306805	140639	133341	141782	145345	74447

Notes: Both Panels report OLS estimates. The unit of analysis in Panel A is a group in Berezkin's catalogue, defined first by the ethnicity of the respondent, then by the language spoken at home and if both are missing by the language of the interview. In Panel B the unit of analysis is an individual. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Individual-level controls include: a quadratic term on age, sex, educational attainment FE, religious denomination FE. In Column (6) we account for the level of generalized trust at the group and individual level, respectively (estimate not shown). Robust standard errors are reported in parentheses in Panel A (clustered at the oral tradition level in Panel B). ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition.

Appendix Table 1 - Panel A: Telugu-related Sources

Author	Year	Source
Elwin, Verrier	1949	Myths of Middle India. Madras: Oxford University Press. 532 p.
Payyand, Raghavan	2006	"Religion – native and alien: Interaction, assimilation and annihilation – a study based on worldview". Discourse of Ideology, Religion and Worldview, M.D. Muthukumaraswamy, ed. Chennai: National Folklore Support Center. P. 198-208.
Zograf, Georgy Alexandrovich	1964	Tales of the peoples of India. Translations from Marathi, Punjabi, Tamil, Telugu, Hindi. Preface and notes G. Zograf. M.-L.
Venkataswami, M.N.	1923	Heerámma and Venkataswami or Folktales from India. Madras: Diocesan Press. 230 p.
Hiebert, Paul G.	1971	Konduru. Structure and Integration in a South Indian Village. Minneapolis: University of Minneapolis Press. 192 p.
Pantulu, G.R. Subramiah	1905	Folk-lore of the Telugus, a collection of forty-two highly amusing and instructive tales. Madras: G.A. Natesan & Co. 139 p.
Grierson, George A.	1908	Linguistic Survey of India. Collected and edited by G.A. Grierson. Vol. 9. Part 2. Specimens of Rajastani and Gujarati. Calcutta: Superintendent Government Printing. 477 p.
Jason, Heda	1989	Types of Indic Oral Tales. Supplement. Helsinki: Finnish Academy of Sciences. 100 pp (FF Communications 242).
Kudinova, M.V., A.M. Kudinov	1995	When luck smiles: Indian tales, legends and folk tales. Comp. and trans. from English Mv Kudinova and A.M. Kudinov. M .: Oriental literature, School Press. 320 s. {According to Folk Tales of India (21 Volumes). Sterling Publishers Private Lmt., 1969-1974}.

Appendix Table 1 - Panel B: The Telugu Oral Tradition

Group	ID	Title
Telugu	a12	Eclipses: a monster's attack
Telugu	a12f	Eclipses: creditor
Telugu	b2a	The female earth
Telugu	b3b	Earth grows big
Telugu	c8	The primeval couple of siblings
Telugu	f100	Test of chastity (a queen and a servant girl)
Telugu	h36	The muddled message
Telugu	h36g	Muddled message: how many meals a day
Telugu	h36g1	Bull is a failed messenger
Telugu	h49b	The faithful dog as security for a debt
Telugu	h7	The personified Death
Telugu	h7c	Not finished prayer
Telugu	i127	Ursa major is a bed
Telugu	i40	Rainbow bow
Telugu	j23	A late son kills monsters
Telugu	k101	Night dances of girls
Telugu	k101c	In the palace by day, on the sky by night
Telugu	k116b	The lecherous holy man and the maiden in the box
Telugu	k129	The disenchanted beauty
Telugu	k130a	Girl in the house of several brothers
Telugu	k136c	Killed and revived periodically
Telugu	k144	The predestined death because of an animal
Telugu	k24	Stolen clothes of supernatural woman
Telugu	k25	Woman from sky-world marries mortal man
Telugu	k27n1	Task-giver is a king or a chief
Telugu	k27nn	Envious minister
Telugu	k27x2	To steal an egg from under the bird
Telugu	k27z	Game of chance for life and death
Telugu	k27z3	Cat with a lamp
Telugu	k27z4	The trained animal of the gambler
Telugu	k37	Recognition-test
Telugu	k38b	Snake threatens nestlings
Telugu	k38c	Bird carries hero to his destination
Telugu	k38e	Of copper, of silver, of gold
Telugu	k56b	The worthy man is rewarded, the unworthy punished
Telugu	k77a	Small objects and animals defeat the ogre
Telugu	k77c	Ones who hide in a house frighten dangerous enemy
Telugu	k82	Evil sister-in-law
Telugu	k83	The sons on a quest for a wonderful remedy for their father
Telugu	k92a	The princess responsible for her own fortune
Telugu	k99b	Eloping with the wrong man
Telugu	l19b	Multi-headed being
Telugu	m114j	All women are similar
Telugu	m130a	Bird helps animal to escape from snare
Telugu	m145	The lion in a well

Appendix Table 1 - Panel B: The Telugu Oral Tradition cont'd

Group	ID	Description
Telugu	m152	Why only one wolf?
Telugu	m152a	Animal tied to another for safety
Telugu	m170	Pilgrimage of the animals
Telugu	m197d	The shortened stick
Telugu	m29b	Trickster-fox, jackal or coyote
Telugu	m29w3	The lion is a failure
Telugu	m39e2	The speaking tree
Telugu	m91c2	Put into bag

Appendix Table 2: Folklore, Institutions and Exchange

	Ln(1+# Motifs on Trade)			Ln(1+# Motifs on Money)			First PCA of Motifs on Trade and Money		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance to Pre-600AD Trade Routes	-0.1637***	-0.2436***	-0.2663***	-0.1610***	-0.1696***	-0.1707***	-0.2740***	-0.3617***	-0.3857***
	(0.0290)	(0.0542)	(0.0556)	(0.0539)	(0.0382)	(0.0453)	(0.0530)	(0.0699)	(0.0814)
Change in Distance to Trade Routes			-0.1134			-0.0054			-0.12
Between 600 AD and 1700AD			(0.0996)			(0.0543)			(0.1198)
Baseline Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Continental FE	yes	no	no	yes	no	no	yes	no	no
Country FE	no	yes	yes	no	yes	yes	no	yes	yes
R^2	0.6	0.596	0.597	0.83	0.853	0.852	0.765	0.776	0.776
# of Observations	474	474	474	474	474	474	474	474	474

Notes: This table reports OLS estimates. The unit is a group in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. All columns focus on groups in the Old World. Standard errors are clustered at the language family as classified by Berezkin. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition.

Appendix Table 3: Pastoralism and Motifs on Violence

	ln(1+# of Motifs on Retaliation (ConceptNet))				ln(1+# of Motifs on Anger (LIWC))			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Predominantly Pastoral	0.1492***	0.0207	0.1430***	0.0716**	0.0641**	0.0617**	0.0317	0.0578***
Groups	(0.0281)	(0.0244)	(0.0491)	(0.0330)	(0.0261)	(0.0272)	(0.0332)	(0.0146)
Baseline Controls	yes	yes	yes	yes	yes	yes	yes	yes
Additional Controls	no	yes	no	yes	no	yes	no	yes
Continental FE	yes	yes	no	no	yes	yes	no	no
Country FE	no	no	yes	yes	no	no	yes	yes
# of Observations	1237	1237	1237	1237	1237	1237	1237	1237

Notes: This table reports OLS estimates. The additional controls are selected by the "post-double selection" methodology of Belloni et al (2015) from the set of 52 BoW in the LIWC and MFD dictionaries (the BoW on anger is naturally not considered). In the "post-double selection" specifications the baseline controls as well as the respective FE are partialled out first. The unit of analysis is a group in the Ethnographic Atlas. A group in the EA has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variables' definitions.

Appendix Table 4: States and the Exchange-Economy

	ln(1 + # of Trade-Related Motifs)						ln(1 + # of Money-Related Motifs)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Jurisdictional Hierarhy Beyond the Village	0.0733*** (0.0213)	0.0671*** (0.0208)	0.0584** (0.0236)	0.0731*** (0.0260)	0.0713*** (0.0234)	0.0677*** (0.0247)	0.1185*** (0.0328)	0.0942*** (0.0252)	0.0889*** (0.0219)	0.1056*** (0.0292)	0.0946*** (0.0248)	0.0898*** (0.0231)
Share of Subsistence from Animal Husbandry			0.0391*** (0.0116)			0.0458*** (0.0127)			0.0242** (0.0116)			0.0178 (0.0128)
Share of Subsistence from Agriculture			-0.0101 (0.0066)			-0.0099 (0.0083)			0.0075 (0.0097)			0.0046 (0.0114)
Baseline Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Additional Controls	no	yes	yes	no	yes	yes	no	yes	yes	no	yes	yes
Continental FE	yes	yes	yes	no	no	no	yes	yes	yes	no	no	no
Country FE	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes
# of Observations	1108	1108	1108	1108	1108	1108	1108	1108	1108	1108	1108	1108

Notes: This table reports OLS estimates. The additional controls are selected by the "post-double selection" methodology of Belloni et al (2015) from the set of 52 BoW in the LIWC and MFD dictionaries (the BoW on money is naturally not considered). In the "post-double selection" specifications the baseline controls as well as the respective FE are partialled out first. The unit of analysis is a group in the Ethnographic Atlas. A group in the EA has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition.

Appendix Table 5: States and Rule Following

	ln(1+ # of Respect-Related Motifs (ConceptNet))				ln(1+ # of Respect for Authority-Related Motifs (MFD))			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Jurisdictional Hierarhy Beyond the Village	0.1100***	0.0716***	0.0855***	0.0768***	0.0689***	0.0309**	0.0289*	0.0214**
	(0.0317)	(0.0240)	(0.0224)	(0.0180)	(0.0203)	(0.0157)	(0.0154)	(0.0104)
Baseline Controls	yes	yes	yes	yes	yes	yes	yes	yes
Additional Controls	no	yes	yes	no	yes	yes	no	yes
Continental FE	yes	yes	yes	no	no	no	yes	yes
Country FE	no	no	no	yes	yes	yes	no	no
# of Observations	1108	1108	1108	1108	1108	1108	1108	1108

Notes: This table reports OLS estimates. The additional controls are selected by the "post-double selection" methodology of Belloni et al (2015) from the set of 52 BoW in the LIWC and MFD dictionaries (the BoW on respect for authority is naturally not considered). In the "post-double selection" specifications the baseline controls as well as the respective FE are partialled out first. The unit of analysis is a group in the Ethnographic Atlas. A group in the EA has been matched to an oral tradition in Berezkin's catalogue. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Standard errors are clustered at the language family level as classified in the EA, v98. ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition.

Appendix Table 6: Oral Traditions and Current Attitudes

Panel A: Across Groups

Group Mean to the Question:	A Respondent finds Justifiable: Claiming Govt Benefits (F114)		Avoiding a Fare on Public Transport (F115)		Cheating on Taxes (F116)		Someone Accepting a Bribe (F117)		Violence Against Other People (F114_03)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln(# Respect-Tagged Motifs from ConceptNet)	-0.4451***		-0.2841		-0.2999*		-0.5712***			
	(0.1612)		(0.1732)		(0.1679)		(0.1594)			
Ln(# Respect for Auhority-Tagged Motifs from MFD)		-0.4183**		-0.5815**		-0.3561		-0.5052**		
		(0.2095)		(0.2288)		(0.2518)		(0.2400)		
Ln(# Anger-Tagged Motifs from LIWC)									-0.1942*	
									(0.1163)	
Ln(# Retaliation-Tagged Motifs from ConceptNet)										0.0935***
										(0.0300)
Baseline Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.113	0.085	0.078	0.096	0.029	0.022	0.148	0.092	0.14	0.177
# of Observations	163	163	164	164	163	163	165	165	84	84

Panel B: Across Individuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ln(# Respect-Tagged Motifs from ConceptNet)	-0.5228**		-0.3451*		-0.3904**		-0.4720**			
	(0.2050)		(0.1923)		(0.1893)		(0.1915)			
Ln(# Respect for Auhority-Tagged Motifs from MFD)		-0.7099***		-0.5759***		-0.4194**		-0.5697***		
		(0.1858)		(0.1798)		(0.1676)		(0.1732)		
Ln(# Anger-Tagged Motifs from LIWC)									0.0536	
									(0.0835)	
Ln(# Retaliation-Tagged Motifs from ConceptNet)										0.0324*
										(0.0176)
Baseline Controls, Country FE, Individual-Level Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.094	0.094	0.103	0.104	0.085	0.085	0.089	0.089	0.105	0.105
# of Observations	296094	296094	295484	295484	295542	295542	306805	306805	74447	74447

Notes: Both Panels report OLS estimates. The unit of analysis in Panel A is a group in Berezkin's catalogue, defined first by the ethnicity of the respondent, then by the language spoken at home and if both are missing by the language of the interview. In Panel B the unit of analysis is an individual. Baseline Controls include: ln(# Motifs), ln(Mean # Words per Motif), # of Publications, and Year of First Publication. Individual-level controls include: a quadratic term on age, sex, educational attainment FE, religious denomination FE. Robust standard errors are reported in parentheses in Panel A (clustered at the oral tradition level in Panel B). ***, **, * denote significance at 1%, 5% and 10%, respectively. See Main Text for variable definition.