Folklore*

Stelios Michalopoulos
Brown University, CEPR and NBER

Melanie Meng Xue
Northwestern University

September 1, 2017

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, culture and its transmission. First, we introduce a unique dataset 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 the group-specific intensity of various traits related to its natural environment, institutional framework and mode of subsistence. We establish that such measures are in accordance with the ethnographic record suggesting the usefulness of folklore in quantifying currently non-extant characteristics of pre-industrial societies including the importance of trade. Third, we use folklore to test various influential conjectures among anthropologists including the culture of honor among pastoralists, the role of women in plough-using groups and the intensity of rule-following norms in centralized societies. Finally, we explore how cultural norms inferred via text analysis of oral traditions predict contemporary attitudes and beliefs.

Keywords: Folklore, Culture, Development, Ethnicity, History.

JEL Numbers. N00, N9, O10, O43, O55

*We are extremely grateful to Yuri Berezkin for generously sharing his lifetime work on folklore classification and providing insightful comments. We would like to thank seminar participants at Tel Aviv university for useful comments and suggestions. Rohit Chaparala and Masahiro Kubo provided superlative research assistance. Stelios Michalopoulos. Brown University, Department of Economics, 64 Waterman Street, Robinson Hall, Providence RI, 02912, United States; smichalo@brown.edu. Melanie Meng Xue. Northwestern, Economics Department; xue5meng@gmail.com.

1 Introduction

Over the last two decades a burgeoning body of work has emerged shedding light on the deep roots of comparative development.[1] This investigation has been greatly aided by moving the empirical explorations at the subnational level and recognizing the crucial importance of groups (ethnic, linguistic and religious) for understanding 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 among historians, anthropologists, geographers, and evolutionary biologists regarding preference formation, institutional and societal traits, beliefs and attitudes, and their consequences for contemporary economic performance, seeDiamond and Robinson (2010), Nunn (2012), Spolaore and Wacziarg (2013), Ashraf and Galor (2017), Michalopoulos and Papaioannou (2017*b*).

This renaissance of the new economic history has naturally given rise to a set of critiques. The first one starts from the observation that in order to convincingly invoke persistence (or change) of cultural traits as an explanation of current outcomes one would like to obtain a measure of these characteristics from the same societies during the pre-industrial era in order to be able to make meaningful comparisons. Attempts to address this issue have been made in the context of specific traits and regions but a broad satisfactory answer is missing.[2] Another closely related criticism centers on the fact that although many of the conjectures regard the historical formation of cultural and societal traits they are being tested against current data. A third critique stresses the weaknesses of the ethnographic work of George Peter Murdock (1967), reflected in the Ethnographic Atlas, including the incomplete coverage of certain economic and social aspects.

In this study we show how integrating folklore in our analysis can greatly expand the scope of the questions asked, open a window into a better understanding of 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, 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 economists and political scientists interested in comparative development and culture.

First, we introduce a unique dataset of folklore that codes the presence of thousands of motifs for hundreds of pre-industrial societies. This is the lifetime work of the eminent

[1] See Michalopoulos and Papaioannou (2017*a*) 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.

anthropologist and folklorist Yuri Berezkin. The underlying texts are sampled from more than 6,000 books and articles. The resulting database contains approximately 50,000 abstracts of oral texts from all over the world, with information on the distributions of more than 2,000 motifs from almost 1,000 societies. Berezkin uses the expressions 'folklore', 'mythology' and 'folklore and mythology' indiscriminately to refer to all kinds of traditional stories and tales, long and short, sacred and profane. His catalogue is described in Berezkin (2015*a*) 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. This is any episode or image related to, or described in, narratives in the oral tradition. Here are some examples of motifs: "impossible riddles", "male sun and female moon", "alive being turns into many objects", "eclipses: relations between the sun and the moon", "primeval tree", "the profitable exchange: from a pea to a horse, "mosquitoes let loose", "task-giver is a king or a chief", etc. 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 dataset to those in Murdock's Ethnographic Atlas (EA) effectively adding the oral traditions to the ethnographic record of pre-industrial societies. We then employ a dictionary-based approach to elicit the group-specific intensity of various traits related to the natural environment, the institutional framework and the mode of subsistence. Groups whose folklore has a higher intensity of earthquake-related motifs live closer to earthquake-prone regions, groups closer to the coast have more motifs reflecting subsistence on aquatic sources, groups on fertile homelands exhibit more motifs related to agriculture, and finally those residing closer to pre-600AD trade routes are more likely to have an abundance of exchange-related motifs. Besides establishing that salient elements of the natural environment are manifested in the oral tradition of the group we also show that folklore-based measures of political complexity and subsistence pattern robustly correspond to the analogous traits in the EA suggesting the usefulness of folklore in quantifying currently non-extant characteristics of pre-industrial societies including the importance of trade.

Third, we test various influential conjectures among anthropologists regarding the cultural and economic attributes of pre-industrial societies including the role of honor among pastoralists, the interplay between political complexity and trade, and role the role of women in plough-using societies. Finally, we explore the association between rule-following motifs and state centralization to shed light on their interaction. We find robust evidence that centralized societies are significantly more likely to have oral traditions espousing rule following, submission to authorities and motifs where the trickster (a common type of motifs) is punished for

his deviant behavior.

Finally, we explore how cultural norms inferred via text analysis of oral traditions translate into contemporary attitudes and beliefs. Specifically, we use the Harvard dictionary and associated General Inquirer Categories for textual content analysis to obtain a host of different folklore-based measures of culture. The latter dictionary dates back to the 1960s and has been widely used in linguistics, psychology, sociology, and anthropology. We demonstrate the predictive power of folklore-based measures of culture on current survey-level outcomes, concluding that folklore itself can 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. In section 3 we provide a brief history of the field of folkloristics and 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 comparative works in mythology, commenting on the timing of folklore. We also introduce and discuss the Harvard dictionary along with the General Inquirer Categories. In Section 4 we detail our empirical approach and present our results in four parts. In section 5 we conclude by offering some thoughts on future work.

2 How Folklore May Help?

Linking pre-WWII economic conditions to current economic performance across countries has been greatly aided by the reconstruction of income per capita series for the currently developed world over the last couple of hundred years, Maddison (2007). For longer time spans cross-country population density estimates by Evedy and Jones (1978) have been invariably used. Similarly, information on institutional variation reflected in the degree of democracy across independent modern states extends back to 1800, thanks to the Polity IV database, greatly facilitating comparisons.

However, group-level historical data are more scarce particularly outside Europe. The only systematic effort to recover the institutional, economic, and societal makeup for a large cross-section of pre-industrial societies is the Ethnographic Atlas. 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. Using a plethora of sources, Murdock (1967) documents a wealth of characteristics mostly for groups in Africa, Asia, Oceania and the New World, prior to contact with Europeans. The results of this major effort are recorded in the Ethnographic Atlas (published in 29 installments in the anthropological journal Ethnology) reflecting societal, institutional, and economic traits

of 1,265 ethnicities.

Thanks to this body of work the research on the cultural, institutional and social correlates of growth has moved beyond country boundaries combining Murdock's EA and the mapping of the spatial distribution of ethnicities, Murdock (1959), with the underlying geographic or location-specific trait to shed light on the origins and consequences of a variety of economic, institutional and cultural traits.[3] Recently, cultural explanations have been thrust in the spotlight, Landes (1998). For example, a recent addition to the list of candidates on the origins of comparative development is a 2015 best-seller by Harari (2015) who makes the bold claim that the success and failure of human societies is deeply-rooted in the common myths that exist in people's collective imagination. Such an intriguing conjecture currently is hard to quantitatively assess. Moreover, cultural explanations have been put forth as driving the observed differences in comparative development, Landes (1998), and there is a heated debate regarding the historical interplay between institutions and culture. Nevertheless, progress in this areas has been hindered by the mere fact that word values surveys are not available for pre-industrial societies. Hence, invoking persistence or change of cultural traits is hard to verify in absence of historical proxies.

This paper proposes a way to recover these traits by showing that folklore can shed important light on a groups' historical heritage including proxies of beliefs and attitudes of ancestral populations conspicuously absent from the historical record. But what is folklore? According to the Oxford 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 (2016). Incidentally but importantly for our purposes, folklore is also an academic discipline whose subject matter (also called folklore) comprises the sum total of traditionally derived and orally or imitatively transmitted literature, material culture, and customs. The insights from this discipline have been so far neglected.

To the best of our knowledge there are no papers in economics that utilize some aspect of folklore. In other social sciences folklore is gradually being integrated. For example, in a recent paper 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

[3]See among other, Nunn and Wantchekon (2011), Michalopoulos (2012), Giuliano and Nunn (2013), Osafo-Kwaako and Robinson (2013), Fenske (2014), Alsan (2015), Fenske (2013), Bentzen, Hariri and Robinson (2015), Mayshar et al. (2015), Michalopoulos, Putterman and Weil (2016), Chiovelli, Cervellati and Esposito (2017), and Michalopoulos, Naghavi and Prarolo (2017).

affiliation exert significant independent effects on folktale diversity.

But how do we analyze folklore? In Section 3 we provide a detailed discussion of the dataset we use. Broadly speaking what we have for each society is a set of motifs (out of a total of 2,320 motifs) indicating the presence of a particular image, episode in the group's oral tradition. A motif comes with a short title and a short usually a two-lines description, e.g., title: "Kind and unkind girls"; Description: "a girl or a woman meets powerful person, behaves herself in a right way and is successful. Another (or two others) behaves in a wrong way and suffers the opposite (is punished or not rewarded)". Note this motif is precisely the one whose diversity within Europe (700 variants) is analyzed by Greenhill and Atkinson (2013) and it is deliberately chosen to illustrate the usefulness of motifs as an aggregator of folktales across multiple variants of a given theme.

This means that our underlying data is text. Over the last few years text analysis has seen great advances and taken a center stage in the social sciences thanks to the abundance of text (from millions of digitized written sources and on-line content). For reviews of text analysis in political science and sociology see Grimmer and Stewart (2013), 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 drawbacks.

The approach we currently employ to quantify folklore is the dictionary-based method which connect counts of specific words to latent, unobserved attributes we wish to quantify. This is the simplest and most commonly-used method. Besides its simplicity we also think that it is appropriate for our setting.[4] 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) and Baker, Bloom and Davis (2016) who use a pre-specified dictionary of terms capturing particular categories of text to obtain an estimate of the outcome of interest. For example, Tetlock (2007) uses bag-of-words specified in the General Inquirer (GI) dictionary to get the sum of the counts of words in each category.[6] In Baker, Bloom and Davis (2016) the authors use the count of articles in a given

[4] We are also currently exploring alternative non-dictionary based methods of text analysis.

[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.

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. The mapping between the two is the raw count of the pre-specified terms divided by the total number of articles in the newspaper-month, averaged across newspapers.

Inspired by these two papers we follow a similar approach. For example, in order to get an estimate of the salience of earthquakes in the folklore of the group we construct the raw count of motifs that mention the word "earthquake". When we want to obtain a measure of the extent to which folklore reflects hostility or aggressiveness we use the corresponding bag of words from the GI and we construct the raw count of motifs that mention any of these words. We always normalize the resulting counts of motifs by the total number of motifs recorded in a given society. It is important to note that while it has only recently been used in economics and finance, the Harvard dictionary and associated General Inquirer categories for textual content analysis dates to the 1960s and has been widely used in linguistics, psychology, sociology, and anthropology. Thanks to its intellectual origins firmly rooted in the social sciences the resulting bag-of-words classifications by the Harvard dictionary are likely to be appropriate for the classification of oral traditions.[7] We return to the issue of cross-validating the Harvard dictionary-based categories below.

An alternative route towards uncovering the cultural background of a given society one may utilize information from the corpus of books published during the pre-industrial era. Such an approach, however, would have to keep into account that book writers during this period and their audience were both very different for the median illiterate person whose beliefs and attitudes we wish to uncover.

3 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.[8] The first generation of folklorists focused exclusively on illiterate peasants, and groups like the gypsies that would be more or less unaffected by the sweeping social changes of the era, and attempted to document

Finally, Wisniewski and Lambe (2013) utilize pre-defined word lists to construct an index of negative-media attention towards the banking sector and find that the former Granger-causes bank stock returns during the 2007–2009 financial crisis.

[7] For example, Loughran and McDonald (2011) show that Tetlock's (2007) approach of using word lists from the Harvard Dictionary is imperfect because it misclassifies common words in financial text and propose an alternative finance-specific dictionary of positive and negative terms.

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

their archaic 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 knowledge and beliefs, 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 pre-dating the spread of Christianity. We return to the issue of the timing of folklore below.

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 accepted limitations of a culture; maintaining cultural identity; and validating existing cultural norms.

For over hundred fifty years from early 19th to mid 20th century a vast body of work accumulated from collectors in all parts of the world listening to story-tellers, and with better and better techniques recording and publishing what they heard. Hence, the very nature of folklore, i.e., 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 (1996). Audiences expected storytellers to retell familiar stories and this 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, i.e., a feedback loop between performing and hearing the performance multiple times, in order to retain the essential elements of the tale, Dorst (2016).

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 motifs lies at the heart of its comparative framework. As the core unit of analysis, motifs can be defined as "the smallest element in a tale having a power to persist in tradition", Thompson (1946). They are hypothetical archetypes established by comparing a large number of texts which share a common core. The methodology most closely associated with the use of motifs in comparative research is the historic-geographic approach that began in the late 19th century. It focused on establishing a folkloric tradition, identifying its geographical starting point and how it spread historically. The method was questioned and criticized in the wake of Post-Modernism, 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 1972, 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-Uther (ATU) index of international folklore which classifies 2,404 motifs, Aarne (2004). The ATU system, however, 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 historical-geographic method by Goldberg (1984). For example, 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. 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.

Berezkin's Folklore and Mythology Catalogue

A critical dimension of the Folklore and Mythology Catalogue is that Berezkin does not limit himself in the European-based ATU motif classification. 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". This assures that the motifs he records are not reflecting only the presence of European motifs among non-European societies. Starting with indigenous societies in the Americas and extending his classification to groups in the Old World over the course of thirty years, Berezkin has compiled a unique database of folklore and mythology for 936 groups worldwide, categorizing more than 50,000 texts into 2,320 motifs. The fruits 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 early pioneers 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), Franz Boas (2002) and his students.[9] By not restricting himself to the extant ATU motifs, he is able to accommodate and classify non-European oral traditions. This ensures that what is explored are potential similarities among oral traditions themselves, rather than dissimilarities between European and non-European folklore. In addition, Berezkin summarizes and makes available the textual sources underlying the motifs, which allows other researchers to verify and interpret the original sources, or to even analyze this wealth of oral traditions independently.

To encode his motifs he has consulted an impressive list of 12,228 books and journal articles. The bulk of the materials in the textual catalogue were published in the 20th century,

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

see Table 2 Panel A for a detailed breakdown by the date of publication as well as some of the sources themselves.

The median group in Berezkin has 58 motifs. The group with the largest number of motifs are the Russians and there is only one group with a single recorded motif; the Yeyi in Nortwestern Botswana. In Table 2 Panel B we report the top-10 groups in terms of the number of motifs whereas in Panel C we report the top-ten 10 motifs in terms of the number of groups they are present. The most popular motifs present in 346 out of the 936 societies in Berezkin's catalogue is the following: "In episodes related to deception, absurd, obscene or anti-social behavior the protagonist is fox, jackal or coyote". In Figure 2 we portray the number of motifs per society in Berezkin's dataset.

Among the $2,320$ motifs, several motif groups emerge: (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, dream, 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 object. (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 (l.) Proper names. (m.) Formulae. Large motif groups such as "Adventures" and "Tricks and Competition" have up to 826 motifs (420 motifs in "Tricks and Competition").

Caveats

For the analysis that follows it is important to keep in mind the following issues.

First, what is the time period the motifs and underlying myths and tales correspond to? The motifs provide a snapshot of folk life from the pre-industrial times, since folklorists were mainly interested in collecting oral traditions from the groups relatively untouched by the waves of modernization of the 19th and 20th century. This data therefore is likely to be a depository of the beliefs and attitudes of the pre-industrial societies. 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 are likely to 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 papers looking at the areal distribution of individuals motifs in relationship to large-scale population movements, migrations, cultural contacts and interactions in history and pre-history, see Berezkin and Duvakin (2016), Berezkin (2015*b*) among others.[10] Hence in the comparisons below with the Ethnographic Atlas it is useful to keep in mind that although folklorists and ethnographers surveyed these societies roughly around the same time period the information coded in folklore is potentially mapping into a longer historical horizon.

Second, Berezkin in his coding of folklore ignores motifs which are 100% universal or only found in a single oral tradition. This is also the case for the ATU classification and it is not surprising since the focus of comparative mythology is on motifs that can be found in different societies and hence are not culture specific. To the extent that both the observed motifs and the unobserved ones (i.e., those that are unique to a given society) are drawn from the same distribution of the themes and images present in the oral tradition then our quantification of group-specific traits is defensible.

Third, we do not have a count of how often a given tale/motif is repeated in the folklore of a society, i.e., we do not know the popularity of a motif, neither the number of variants of particular tale (which can be numerous as the study by Greenhill and Atkinson (2013) reviewed above suggests). Hence, our folklore values reflect the extensive margin of elements in the oral tradition. 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 linguistic lines but sometimes he groups related societies together when he cannot find enough information on the oral tradition of each individual language. For example, he puts together the neighboring groups of the Fulbe, the Wolof and the Serer located in Western Africa.

Interpretation of Folklore

In the early years of folklore, the main task of folklorists has been to collect and classify different folklore materials paying relatively little attention to its interpretation. Indeed, the primary focus has been on the "lore" per se. For our purpose, however, it is important to understand

[10] Other famous scholars like Witzel (2012) contend that the earliest mythology, the Gondwana one, dates back to 65,000 years ago, see [p.416] of his book.

the relationship between the "folk" and the "lore", i.e., it is not enough to say that folklore is a mirror of a culture. To understand the meaning of folklore, we need an approach (or several) to operationalize the process of information extraction. There are three dominant approaches: the humanistic, the anthropological and the psychological one. The latter two are particularly relevant in our context.

According to the pattern theory of culture (see anthropologist Ruth Benedict), all parts of culture are related and reflect the same values and beliefs. Based on her theory, folklore can be seen as a window into culture. 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). Within psychology, 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, Gustav (1968). They consider the distance past as "hidden in the unconscious and reflected in folkloric symbols", Green (1997).

Because we work with motifs and summaries of myths, legends and tales - both of which are intentionally deprived of details - the humanistic approach which mainly emphasizes the role of the narrator, offers few insights. Between the anthropological and psychological approach, we do not discriminate between the two. We take a combined approach to maximize the amount of information we may extract from the body of folklore materials. Specifically, we employ the Harvard General Inquirer, which is known to be a useful tool for both psychoanalysts and cultural anthropologists.[11]

Harvard General Inquirer

The Harvard General Inquirer is a lexicon attaching syntactic, semantic, and pragmatic information to part-of-speech tagged words, Stone, Dunphy and Smith (1966). In 1962, the General Inquirer was first developed to research problems in the behavioral sciences. It was part of an attempt to develop more formalized procedures involving non-numerical data. Its aim was to produce a method for automatic theme analysis. The current version of the General Inquirer comprises (a) the Harvard IV-4 dictionary, (b) the Lasswell value dictionary, (c) several categories recently constructed, and (d) "marker" categories. Altogether, the General Inquirer has 182 categories with each category having a range of 6 to $2,045$ words.

Taking the Lasswell dictionary as an example, the dictionary's entries were developed by Namenwirth and Weber (2016). Their book *Dynamics of Culture*, originally published in 1987, is "a landmark contribution to macrosociology that extends the tradition of Sorokin, Durkheim,

[11] http://www.wjh.harvard.edu/~inquirer/3JMoreInfo.html

Marx, Weber and other founders of the discipline". In the book, they discuss their research on culture indicators over two decades' time. The Lasswell dictionary provides a list of words in four reference domains: power, rectitude, respect, affiliation, and four welfare domains: wealth, well-being, enlightenment and skill. Within each domain, there are subcategories reflecting gains, losses, participants, ends, and arenas. Thus there are words associated with power increasing (*power gain*) and words associated with power loss; rectitude gain and rectitude loss as well as religion and ethics.

From a practical point of view, the General Inquirer is a mapping tool. It maps a given text to counts of dictionary-supplied categories. Many of these categories may potentially capture meaningful cultural indicators ("submission", "persistence", "status and prestige", "hostility and aggression"), etc. To use this tool, we first break down each motif description into words. Then we look up all the words appearing in the description and tag each one to the corresponding category(ies). Hence, for each motif description we have a binary variable of 1 or 0 per motif, per tag. As a last step and to arrive to our group-specific estimates we add up all motifs within each specific dictionary category to quantify the intensity of a particular cultural indicator within that oral tradition.

Our dictionary-based approach provides an initial examination of the cultural context of folklore. It offers several advantages including the minimization of subjective interpretation, i.e., the use of ad-hoc bag of words (which as it will become evident we also do in some parts of the analysis). By focusing on the motifs and summarized texts, we only make use of the essential plot of a story. This largely removes subjective influences imposed by the narrator. The objectivity of our analysis is completed by the application of a dictionary-based approach, which imposes strict discipline to the interpretative process. However, the objective nature of this approach also entails its limitations. The General Inquirer 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."[12] In addition, the General Inquirer was created in the early 1960s. A certain proportion of words classified by the General Inquirer applies to the modern context. Given the antiquity of folklore, it is possible that aspects of oral tradition remain unexplored when viewed through the lens of the General Inquirer.

4 Empirical Analysis

We conduct four different exercises to gauge the predictive power of folklore regarding an array of different group-specific traits. These include (i) whether a group's physical environment can

[12] http://www.wjh.harvard.edu/~inquirer/3JMoreInfo.html

be discerned in its folklore, (ii) how folklore-based measures of subsistence and institutions relate to the ethnographic record, (iii) the potential of folklore to elucidate famous conjectures among anthropologists regarding preference formation and exchange, and finally (iv) how the pre-industrial oral tradition of a group relates to contemporary beliefs and attitudes of its members.

4.1 Folklore and the Natural Environment

In this section we ask whether salient features of the physical environment are reflected in the oral tradition of the society. This first step serves the purpose to check the extent to which the natural environment leaves an imprint on the folklore images and episodes of a group. This is not a trivial exercise. Berezkin's view of folklore as a depository of the migration history of a society suggests that folklore elements can be preserved even if the landscapes, climates and social configurations in which they are told change, Berezkin (2015*a*). With this in mind we check the following six traits which we think we can reliably measure both in the folklore and in the physical environment of a group. Specifically, we look at features of the landscape that presumably have not changed dramatically over the course of modernization. These include proximity to the coast, proximity to earthquake zones, intensity of lightning strikes, malaria ecology, and caloric suitability for agriculture for the crops available in pre-1500 CE. See Table 1 Panel *A* for the respective summary statistics and correlation matrix. Are these natural phenomena salient/important enough to manifest themselves in the oral tradition of a society? This is what we explore below.

The dependent variable in Panels *A* and *B* of Table 3 is a count variable so we adopt the following specification and run Poisson regressions.[13]

$$Topic\text{-}Specific\ Motifs_{i,c} = a_c + \beta GEO_i + \gamma \ln(Total\ Motifs_i) + \varepsilon_i$$

$Topic\text{-}Specific\ Motifs_i$ is the number of motifs reflecting a particular characteristic of group i. GEO_i is vector of geographic traits. We use the group's centroid (recorded by Berezkin) to compute the distance terms and a radius of 250 kilometers around each group to get the values of the respective geographic and ecological trait. The number of motifs per group in principle is a very interesting variable itself partially reflecting the underlying breadth of themes, images and episodes present in the oral tradition. However, the same variable also naturally reflects both the intensity at which a given society has been sampled by folklorists and eventually Berezkin himself. Since we cannot distinguish between the two, we will always

[13] Results are similar using OLS or the logged versions of the respective motif categories.

control for the log number of motifs per group.[14] When we use only Berezkin's database we cluster the standard errors at the level of the language family as recorded by Berezkin himself. When we combine the former with the Ethnographic Atlas we cluster the standard errors along the language affiliation coded by Murdock (v98).

a_c reflects continental or country-specific constants. Why do we control for country fixed effects 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 pre-industrial historical legacies across modern-day countries, the inclusion of country fixed effects is primarily motivated by concerns about Berezkin's sampling of folklore from societies in different parts of the globe. One may worry, for example, that the coverage is systematically poorer in certain parts of Asia or Melanesia compared to parts of Europe or the Americas. Hence, by including country-specific constants we mitigate concerns of potentially unbalanced coverage across countries and also in the quality and breadth of the underlying recorded oral traditions (assuming that groups within the same country are likely to be sampled by folklorists during the same time period and presumably with similar biases/available technology). Hence, exploiting within-country cross-group variation increases our confidence that the uncovered patterns are not an artifact of variation in folklore sources across modern-day countries.

Images and Episodes in Folklore Reflecting the Physical Environment In columns 1 and 2 of Table 3 Panel *A* the dependent variable is the count of motifs that mention the word "earthquake". There is a total of 6 motifs. Invariably these motifs offer a rationale for why earthquakes occur such as: "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", or "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 located in zones 3 and 4). An average group in Berezkin's dataset has 0.10 earthquake-related motifs. However, those located within an earthquake zone (that is, have a distance of 0) have on average 0.20 such motifs twice as many compared to groups located outside these areas (mean 0.09). In columns 1 and 2 we show that this pattern is robust to accounting for continental and country fixed effects.

In columns 3 and 4 we count motifs that mention the words "thunder", "lightning", "storm", "cloud", "rainbow", "deluge", "flood", "cataclysm", and "rain". To measure the intensity of these phenomenona we use the Gridded climatologies of the mean lightning flash rates

[14] We have also experimented flexibly controlling for the number of motifs per group adding decile-fixed effects. The patterns are unchanged.

observed by the spaceborne Optical Transient Detector (OTD) and Lightning Imaging Sensor (LIS) instruments from 1995 to 2010, see Cecil, Buechler and Blakeslee (2014). Pre-industrial societies located in regions experiencing intense thunder strikes systematically feature images and episodes in their oral tradition mentioning instances of these meteorological phenomena,

Finally, in the last two columns we ask whether the disease environment reflected in the intensity of malaria transmission influences the presence of motifs mentioning one of the following words: "mosquito", "insect", and "worm". The data on malaria stability come from Kiszewski et al. (2004). Groups in high-malaria regions are significantly more likely to feature tales and legends where mosquitoes and worms play a prominent role. Here is a representative motif present in 81 societies (the most widespread of this type): "Stinging insects (rare diseases) had been inside a container or some enclosure. They escaped to the world when the container or enclosure was foolishly opened".

Mode of Subsistence Inferred from Folklore and the Physical Environment It is well understood that the environment exerts a significant influence on the mode of subsistence a group will undertake. For example, groups residing on more fertile lands are on average more likely to depend on agriculture for subsistence, whereas those located closer to the coast may naturally incorporate in their diet a wide set of aquatic sources. In Panel B of table 3 we ask whether this relationship is evident also in the oral tradition of a group. Coastal proximity is measured from the centroid of each group whereas we use the caloric suitability of agriculture in a given region for crops available before the Columbian exchange, i.e., before 1500 AD using the data from Galor and Ozak (2015). But how do we get a proxy of the importance of agricultural and fishing activities from a society's folklore?

We constructed two ad-hoc bag of words choosing those elements that we feel are clearly related to the respective mode of subsistence. For the agricultural activities these are the words we took into account: "bread", "grain", "cereal", "agriculture", "farm", "seed", "field", "harvest", "cultivate", "manioc", "wheat", "crop", "plow", and "rice". An average group in Berezkin has 1.52 motifs with at least one word from the list above. In the first two columns the dependent variable is the number of agriculture-related motifs and the dependent variable of interest is the log(mean caloric suitability pre-1500 AD). Among groups that have 0 farming-related motifs the average regional caloric suitability for agriculture is $1,087$ whereas the caloric suitability jumps 30% among groups with at least one farming-related motif reaching $1,416$. This pattern is the same exploiting within-continent variation and within-country variation.

To capture the intensity of fishing in the oral tradition of a group we counted the motifs that had at least one word from the following list: "fish", "canoe", "boat", "harpoon", "hook",

"net", "sea/water mammal". In the oral tradition of an average group there are 2.59 fishing-related motifs. The median group in Berezkin is within 201 kilometers from the coast. Among these groups the number of fishing-related motifs is 2.83 about 20% more than the respective number (2.36) among groups located far from the coast. In columns 3 and 4 of Panel we show that this link is strong both within continents and within-modern day countries (see Figure 3 for the spatial distribution of fishing-related motifs normalized by the total number of motifs). Overall the results in these two Panels provide strong evidence on the influence of the geography and ecology on the folklore-based measures of subsistence and of images in the oral tradition reflecting distinct features of the geographic landscape. But are the these folklore-based measures of subsistence predictive of the actual lifeways as recorded by ethnographers? We answer this question in the next section.

4.2 Folklore and the Ethnographic Record

To explore the correspondence between the oral tradition of a group and its ethnographic record we linked the societies in Berezkin database to those in the EA. Specifically, out of the $1,265$ groups in the EA we found a corresponding group in Berezkin for $1,233$ of these societies resulting in a match rate of roughly 98%. From the 936 groups in Berezkin we matched 613 to these 1233 in the EA implying no ethnographic coverage for approximately one third of societies for which Berezkin has systematized its folklore. Generally, we run OLS specifications of the following type:

$$EAtrait_{i,c} = a_c + \beta \ln(1 + Topic - Specific\ Motifs_i) + \gamma \ln(Total\ Motifs_i) + \varepsilon_i$$

where $EAtrait_i$ is the trait of interest from the Ethnographic Atlas and $\ln(Topic\text{-}Specific\ Motifs_i)$ is the log number of motifs belonging to a given category. a_c are continent/country fixed effects. See Table 1 Panel B for summary statistics and the correlation for this sample.

Mode of Subsistence in the Oral Tradition and in Ethnographic Record

Table 4 Panel A aims at showing that the folklore-based measures of the specific subsistence modes are in accordance with actual measures of economic activity recorded in the EA. Needless to say we interpret these regressions as conditional correlations. Our goal in this part of the paper is to provide sufficient evidence that folklore-based measures (of the economy and the society) are strongly correlated with their observed counterparts from the EA. Our hope is that by doing so we can then use folklore to deduce other aspects of a group that are plainly absent from the EA, like the degree of the market/exchange economy.[15] Hence, we view oral

[15] To get at missing characteristics from the EA the Standard Cross Cultural Sample (SCCS) is an immensely detailed ethnographic dataset. Its drawback is its limited coverage across societies.

tradition and the ethnographic record as providing (noisy but as to be shown below correlated) information for the same underlying social, economic and institutional structure.

Our main regressors in the first 4 columns of Table 4 Panel *A* are the ln(1 + number of farming-related motifs), the ln(1 + number of hunting, gathering and fishing (hgf)-related motifs) and the ln(1 + number of herding-related motifs). This classification of motifs boils down to determining a bag-of-words representative of the activity considered. Above we already discussed how we classify the agriculture-based motifs. Regarding the hgf motifs the words we use are: "hunt", "ungulate", "sledge", "sleigh", "boat", "stag", "opossum", "elk", "buffalo", "deer", "gazelle", "antelope", "bow", "game", "mammoth", "harpoon", "hook", "arrow", "net", "mammal". With respect to the pastoral motifs this is the list of words we consider: "cattle", "goat", "cow", "camel", "horse", "graze", "lamb", "herd", "shepherd", and "pasture". In columns $5-8$ we further break down the hgf motifs into fishing-specific ones as defined above, and hunting-specific ones focusing on the following subset of words: "hunt", "ungulate", "sledge", "sleigh", "stag", "opossum", "elk", "buffalo", "deer", "gazelle", "antelope", "bow", "game", "mammoth", "arrow", and "mammal".

The dependent variable 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 and hunting in columns 5-6 and 7-8, respectively. 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 both when we exploit variation between and within countries. For example, groups with more farming motifs in their oral traditions are systematically more likely to be farmers and those with motifs that describe hgf activities less likely so. Interestingly, for pastoral societies both herding and farming-related motifs are predictive of their dependence on herding with the beta coefficients estimates in column 3 being three times as large for pastoral motifs (0.33) compared to farming ones (0.11). Finally, hunting-specific motifs are robust features of hunting societies, and fishing-only motifs are key elements of the oral tradition among societies characterized by a greater reliance on aquatic resources. An alternative way to gauge the predictive power of folklore-based measures of subsistence and observed modes of subsistence is to look at the partial R^2. This is between 0.20 and 0.35 in these regressions suggesting that a significant fraction of observed variation in a given mode of subsistence can be accounted for by variation in the intensity of the respective motifs.

Institutions and Exchange in Folklore, the Ethnographic and the Historical Record The first two columns of Table 4 Panel *B* follow a strategy similar to Panel *A*.

Specifically, we ask whether folklore-based measures of political complexity predict the strength of observed political complexity in the EA. So, we are building the following bag-of-words to capture the presence of hierarchy in the society: "leader", "hierarchy", "chief", "prince", "queen", "sovereign", "noble", "elite", "ruler", classifying motifs accordingly. An average society in the EA (with a folklore from Berezkin) has 1.71 motifs reflecting some presence of institutional complexity. This number, nevertheless, is 3.84 for centralized societies and 0.98 for non-centralized ones.[16] Columns 1 and 2 highlight that the pattern found in the simple summary statistics is present within continents and countries. More hierarchical groups in the pre-colonial era as recorded in the EA are systematically more likely to have motifs that reflect this hierarchy.

In the last 3 columns of Table 4 Panel B we ask a different question. Motivated by the finding that folklore does seem to reflect features of the observed economic and institutional structure of a group we use folklore to obtain a measure of the intensity of the exchange economy in that society. To do this we construct a bag-of-words that aims at capturing the presence of trade and exchange. Specifically, we categorize a motif as belonging to the trade category if it contains at least one of the following words: "exchange", "buy", "creditor", "lender", "sell", "sold", "bazaar", "trade", "caravan", "price", "money", "coin", and "market". The average group in the EA group has 1.56 motifs related to exchange (see Figure 5 for the global distribution of such motifs). Is there a way to verify whether the observed 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 historical trade routes to the observed intensity of exchange-related motifs.

In other words, is it the case that groups closer to the pre-industrial trade routes are likely to have trade-related motifs in their oral tradition? To shed light on we used data from Michalopoulos, Naghavi and Prarolo (2017) that put together for the Old World a comprehensive set of pre-600 AD 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 the EA to the nearest pre-600 AD route. The summary statistics are telling of a robust broad pattern. Among the 774 societies in EA located in the Old World those within 100 kilometers of ancient trade routes have an average of 6.23 exchange-related motifs, a number 3 times as large compared to groups located further than 100 kilometers (which have only 1.28 of such motifs). Columns 3 and 4 show that pattern is not driven by broad differences

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

across continents or modern-day countries highlighting the usefulness of folklore in quantifying missing important aspects of pre-industrial societies. In column 5 we add a control for distance to trade routes as of 1800 AD. Interestingly, the latter is insignificant whereas the pre-600 AD coefficient remains precisely estimated. This pattern suggests that although we only have a snapshot of the oral traditions across societies around the turn of the 20th century, elements of folklore are likely to encode information on the economy and the society harking back several hundred of years.

4.3 Folklore & Long-Standing Conjectures in the Social Sciences

Having established that information distilled from folklore can complement the ethnographic record below we venture into testing outstanding hypotheses among anthropologists on the formation of attitudes and the state that have received attention also among economists.

4.3.1 State Centralization and the Exchange Economy

The first thing we check is whether historical states are more likely to engage in trade. This has been elevated to an article of faith among economic historians but evidence on the extent of the exchange economy is sparse for pre-industrial societies particularly outside Europe and often based on qualitative accounts. Armed with the measure we constructed above on the intensity of exchange in the oral tradition in Panel *A* of Table 5 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 little empirical counterpart.

On the one end, the median society with no mention of exchange in its oral tradition is a stateless society, i.e., it has no levels of jurisdictional hierarchy above the local, village level. On the other end, groups with at least 3 motifs on exchange have a median of two layers of political complexity in their group. Columns 1 and 2 suggest that this pattern is strong both across and within countries. In column 3 we add two variables to account for the share of subsistence that comes from agriculture and pastoralism, respectively. There are couple of interesting observations. First, political centralization remains a robust correlate of exchange besides the relationship between the mode of subsidence and exchange itself. Second, the intensity of exchange-related motifs for agricultural groups is no different compared to hgf 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 pre-industrial era are not necessarily more likely to engage in trade to the extent that they can

satisfy a large part of their subsistence needs from agricultural products, whereas pastoral ones, given the limited set of resources they produce, would have to rely systematically more on trade and exchange for their survival. Richerson, Mulder and Vila (2001), for example, observes 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."

4.3.2 The Role of Women and the Plough

Boserup (1970) put forward an interesting hypothesis attributing contemporary differences in gender norms to the type of agriculture practiced in the pre-industrial era. The specific hypothesis links the use of the plough to women specializing in home production. The idea is that unlike shifting cultivation the plough requires significant upper body strength favoring male labor. Alesina, Giuliano and Nunn (2013) marshal impressive evidence in favor of this conjecture showing that groups of people originating from regions where the plough was historically used have less equal gender norms today.

One could use the oral tradition to shed further light on this issue. For example, one may attempt to extract from the folklore of a group the number of motifs in which women are depicted favorably to obtain a measure of historical gender bias. We are not pursuing this further here but instead we do something simpler but we hope equally illuminating. Going over Berezkin's catalogue we noticed a particular motif titled the "The epoch of women" with the following description: "The women dominated over the men in the past or in a far away land, were the active part in marriage relations, practiced activities which now are reserved for men only". This motif is present in 64 pre-industrial societies scattered around the globe. To the extent that the adoption of the plough was consequential for the role of women in the society one might expect to see such groups to have this theme in their oral tradition.

According to the EA among the $1,129$ groups, 134 were already using the plough when surveyed by the ethnographers. Among these plough-using groups the probability of having the above-mentioned motif, reflecting a decline in the status of women, is 17% whereas the corresponding number among non-plough societies is three times smaller, namely 6%. Columns 4 and 5 in Table 5 Panel B show that this pattern is robust when we exploit within-continent variation but becomes less precisely estimated but of similar magnitude when comparing groups within countries. We plan to further assess the stability of this pattern by exploiting geographic variation in agricultural suitability similar to Alesina, Giuliano and Nunn (2013).

4.3.3 Culture of Honor in Pastoral Societies

Nisbett and Cohen (1996) conjectured that the high prevalence of homicides in the South of the United States was due to the culture of honor that originates from the settlement by herders from the fringes of Britain in the 18th century. Grosjean (2014) provides robust evidence along these lines. The idea behind this link is that pastoral societies are likely to rely heavily on aggression and male honor as a way to avoid having their herd stolen. In such environments of imperfect property rights and easily movable property aggression and a willingness to kill may create a reputation for toughness and hostility and deter instances of theft. To increase the plausibility of this argument one would like to see whether indeed pastoral societies in the past valued male aggression and hostility. How can one deduce whether a group's oral tradition places a disproportionate emphasis on aggression?

To extract such information from the oral tradition of the group we rely on the categorization from the Harvard Dictionary that features a bag-of-words indicating "an attitude or concern with hostility or aggressiveness". Examples of words that are included are: "aggression", "revenge", "assail", "assault", "attack", "avenge", "beat", "clash", "challenge" etc. Predominantly pastoral groups have on average 35.53 motifs reflecting aggression and hostility whereas the corresponding number among non-pastoral societies is 24.68. Columns 6 and 7 in Table 5 Panel *B* show that this relationship is robust to continent and country fixed effects. This finding is important for two reasons. First, it provides large-scale evidence in favor of an influential (and intensely debated) conjecture about the culture of aggression among pastoral societies showing that the latter feature more prominently in their oral traditions episodes and images that are more violent and aggressive than otherwise. Second, this finding helps also understand how a particular attitude (in this case a proclivity to hostility and aggression) may survive across generations even when the original conditions, that made this type of cultural adaptation optimal, no longer apply. A group's collective memory enshrined in its oral tradition may be this vehicle of cultural transmission.

4.3.4 Rule Following and Political Complexity

Are strong states associated with rule-following norms? From a theoretical standpoint the answer is ex-ante ambiguous and it boils down to whether institutions crowd in or crowd rule-following. From an empirical point of view the existing findings go in both directions. For example, Lowes et al. (2015) compare descendants of a centralized group, the Kuba, to those of 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) focuses on

the Dai Viet-Khmer boundary within Vietnam and finds that strong historical state crowded in village-level collective action and local governance. Can the oral tradition of a group help to shed light on this question?

We construct with 3 alternative measures of rule following in the folklore of each society and show that groups that were politically centralized during the pre-industrial era their oral traditions would be systematically more likely to feature motifs that (i) indicate moral imperative, (ii) a larger frequency of words suggesting submission to authority and (iii) a larger frequency of motifs where anti-social behavior is punished. Although this relationship is not intended to be interpreted as causal is suggests that historically rule-following norms and state centralization on average have been going hand in hand.

The Harvard dictionary has 27 words that indicate moral imperative. Among those the most common in the motifs are words like: have to, must, ought to, should, etc. Here is a description of such a motif: "to return or to get his or her son, wife, husband, domestic animal or (rare) object, person *must* recognize her, him or it among several identical persons, animals or objects". This is a rather common one being present in 138 out of the 936 societies in Berezkin's dataset.

Another category in the Harvard dictionary is the one connoting submission to authority or power, dependence on others, vulnerability to others, or withdrawal and it comprises of a set of 284 words. These include words like: "admit", "accept", "apologize", "commit", "comply", "follow", "serve", "submit", "impossible", etc. Here is a motif, present in 41 societies, classified in this category: "Antagonist names numbers, every time asking what is it. The hero gives answers that the antagonist *accepts* as correct ones." Another motif reads: "A powerful person covets a beautiful bride or wife of a man and gives him *impossible* tasks to get rid of him".

A simple cross-tabulation of the intensity of these categories and state centralization as reflected in the EA is telling of the underlying pattern. On the one hand, stateless groups have on average just 1.29 motifs that indicate moral imperative whereas the respective number for large (complex) centralized states (v33=4 or v33=5) is almost 4 times larger with an average of 4.76 such motifs. The gap is also evident regarding the number of motifs that suggest submission to authority with politically acephalous societies having half the number of such motifs compared to centralized ones (4.91 versus 11.87) . In Table 5 Panel *B* we show that these patterns are robust to adding continental fixed effects in columns 1 and 3 and perhaps more interestingly even comparing groups within the same modern-day-country boundaries, political complexity and rule-following motifs go in tandem.[17]

In the last columns of Table 5 we move away from the Harvard dictionary categories

[17] In results available upon request we show that this correlation is not driven by the presence of high gods in the group.

and we do the following. As we discussed in Section 3 a common type of motifs are those that relate to trickster stories. Closer examination of these motifs reveals that trickster related-stories can roughly be broken down into motifs where (i) the trickster is punished for his anti-social behavior; (ii) the trickster is successful and gets away with his actions and (iii) motifs where we cannot tell what happens to the trickster after all. An example of the first case is the following motif present in 67 societies: "A stranger tells a woman that he comes from the other world and had seen there her dead relative. The woman gives him money and goods for the latter. The husband goes after the trickster to retrieve the money, the trickster steals his horse." An instance of a motif where the cheater/trickster is punished (present in 20 groups) is the following: "Rock chases or otherwise punishes person who has offended it". Finally, motifs like this one (present in 100 groups): "In episodes related to deception, absurd, obscene or anti-social behavior the protagonist is a turtle (or tortoise), a toad or a frog" cannot be classified in either category. It stands to reason that an oral tradition featuring many motifs where the trickster is punished instead of being rewarded is a society with strong rule-following norms.

In order to systematically classify trickster stories into these 3 groups we recruited 3 undergraduates from Brown university that read over each motif description and manually classified a motif into one of the said categories (if any). For each group we aggregated the students' responses into the 3 categories and extracted the first principal component of each. The results are shown in columns 5, 6 and 7. In the first column where we do not distinguish between the different types of trickster motifs we find no systematic relationship between statehood and the frequency of trickster motifs. The pattern changes in column 6 where we distinguish between these 3 categories. Among centralized societies there is systematically higher frequency of motifs where the trickster is punished. The same pattern is found exploiting within-country variation in column 7. All in all the results of Table 5 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.[18]

4.3.5 Folklore and Contemporary Beliefs and Attitudes

Does folklore shape contemporary attitudes? To the extent that folklore reflects the "intellectual remains of earlier cultures surfing in the traditions of peasant class", do such intellectual remains eventually die out or reverberate into the contemporary beliefs and attitudes of people? To answer those questions, we conduct the following exercise: first, we link survey respondents

[18] To increase our confidence in the Harvard dictionary-based results we plan to manually audit the resulting classifications. An illustrative example of this is Baker, Bloom and Davis (2016), who perform a careful manual audit to validate their dictionary-based method for identifying articles that discuss policy uncertainty.

in the WVS-EVS to one or several closely related folklore traditions, using information on their ethnicity, language spoken at home, and language used in the interview.[19] Next, we extract the characteristics of oral traditions as reflected in the distribution of motifs across the 182 categories in the Harvard General Inquirer.

We start our analysis with categories with a precise definition and that can be meaningfully interpreted in the context of WVS-EVS. Figure 6 shows our preliminary findings. Examining traditions with a large number of motifs containing causality-related words, we find that individuals from such traditions are far more likely to say that they value "feeling of responsibility" in their children. This is not surprising given that the knowledge of the consequences of one's actions is a stepping stone for an individual to act responsibly.

We now turn to another important concept: autonomy. In WVS-EVS, the autonomy index takes on values ranging from -2 "Religious faith/obedience" to 2 "Determination, Perseverance/Independence". Are individuals from an oral tradition where many motifs touch upon the concept of submission, using words such as"admit", "conform" and "fear", more likely to be on the left end of the spectrum?[20] And similarly, are individuals from an oral tradition emphasizing perseverance and endurance more likely to develop a determined and independent character? Preliminary shown in Figures 7 and 8 below suggests this to be the case.

The categories provided by the Harvard General Inquirer often do not have a one-on-one relationship to questions in WVS-EVS. The General Inquirer is well-suited for a theme analysis, but we are interested in the meaning of the folklore. To refine our analysis, we begin to classify our texts using a 2×2 metric: tone + theme. That is to say that after we determine the relationship between a motif and a particular category, we apply a sentiment analysis to identity the tone of the text. Crude still, this approach allows us to distinguish between motifs conveying messages going in opposing directions. Figure 9 represents an initial effort in applying this approach.

Related to work there are hundreds of motifs. However, without knowing how work is portrayed in folklore, i.e., positively or negatively, there is no real foundation for studying whether there has been any persistence in attitudes towards it. Using the existing word-count approach to quantify the intensity of work-related motifs in each folklore tradition, we further classify the texts into having a positive versus a negative tone. We find that work, when discussed in a context with predominately negative emotions, reliably predicts negative

[19] When an individual is linked to more than one folklore tradition, we treat a motif as present as long as it is present in any of those folklore traditions this individual belongs to. Besides, when ethnicity information is lacking, we supplement it with the ethnic region an individual belongs to, when ethnic labels overlap with regional identity.

[20] The "submit" category has 284 words, connoting "submission to authority or power, dependence on others, vulnerability to others, or withdrawal".

attitudes towards work today.

Our initial examination of WVS-EVS with the Harvard General Inquirer suggests that images and episodes encapsulated in a group's oral tradition have the tendency to influence the way individuals think about the world and they way they live their life. To get a better idea of the relationship between past folk culture and present value systems, however, we need to better understand our textual sources and refine our methods. In the future, as more extensive surveys on cultural and social values become available for more parts of the world, the role of folklore in connecting the past and the present stands to be further crystallized.

5 Concluding Remarks

The economics on the cultural determinants of growth starting with Landes (1998) have placed a great emphasis on the importance of culture for determining contemporary economic and political outcomes. To overcome the issue of endogeneity very 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, Fernandez (2011)). Nevertheless, there has been a missing element in this literature. Namely, the absence of historical proxies of beliefs has severely hindered the debate about the persistence of cultural traits.[21] 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 dataset of oral tradition across approximately $1,000$ pre-industrial 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, its institutional complexity and extent of the exchange economy. We show that these folklore-based measures are predictive of the observed natural landscape, and the economic and societal features as recorded in the Ethnographic Atlas. Third, we put to test various famous hypotheses among social scientists regarding the relationship between cultural beliefs and economic traits of pre-industrial societies including the role of women in plough societies, the culture of honor among pastoralists, rule-following norms in the oral tradition of centralized groups and the frequency of exchange in the latter. Finally, we explore how certain beliefs and attitudes deduced from the folklore of a group influences the contemporary beliefs and attitudes of its members. For the latter we use the Harvard dictionary and associated General Inquirer Categories which have been widely used in linguistics, psychology, sociology, and anthropology

[21] 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 analyze text.

We view this study as a springboard for further research. For example, one can utilize the wealth of folklore to derive bilateral measures of historical cultural proximity across groups and countries. This is likely to complement the existing bilateral distance measures based on languages, and genes, Spolaore and Wacziarg (2009). Moreover, to the extent that some beliefs and attitudes are more likely to persist than others folklore can shed light on which values are largely stable and which ones are subject to change. An alternative dimension that folklore can be employed is along the lines of Berezkin's work, i.e., using it to trace the historical migration paths of pre-industrial societies. Finally, although obtaining time variation in folklore is challenging granted the inherent uncertainty with timing the origin of a given motif, one may extend this analysis to obtain relatively high-frequency measures of oral traditions using text from contemporary popular culture. Given the versatility of folklore a vehicle of obtaining a unique view of our ancestral cultural heritage we expect it to be widely used among scholars interested in the historical origins of comparative development and culture.

References

Aarne, Antti. 2004. *The Types of International Folktales: a Classification and Bibliography, Based on the System of Antti Aarne and Stith Thompson.* Helsinki, Finland:Suomalainen Tiedeakatemia, Academia Scientiarum Fennica.

Alesina, Alberto F., and Paola Giuliano. 2016. "Culture and Institutions." *Journal of Economic Literature, forthcoming.*

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics*, 128(2): 169–530.

Algan, Yann, and Pierre Cahuc. 2010. "Inherited Trust and Growth." *American Economic Review*, 100: 2060–2092.

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

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

Ashraf, Quamrul, and Oded Galor. 2017. "The Macrogenoeconomics of Comparative Development." Journal of Economic Literature, forthcoming.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2016. "Machine Translation: Mining Text for Social Theory." *Quarterly Journal of Economics*, 131(4): 1593–1636.

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

Bentzen, Jeanet, Jacob Gerner Hariri, and James A. Robinson. 2015. "The Indigenous Roots of Representative Democracy." NBER Working Paper 21193.

Bentzen, Jeanet Sinding. 2015. "Acts of God: Religiosity and Natural Disasters Across Subnational World Districts." University of Copenhagen Working Paper.

Berezkin, Yuri. 2015*a*. "Spatial Methods." In *The Retrospective Methods Network Newsletter*. Vol. 10, , ed. Frog, Helen F. Leslie-Jacobsen and Joseph S. Hopkins, Chapter Between Text and Practice: Mythology, Religion and Research, 58–70. Helsinki:Folklore Studies / Dept. of Philosophy, History, Culture and Art Studies University of Helsinki, Helsinki.

Berezkin, Yuri. 2015*b*. "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.

Berezkin, Yuri, and Evgeny Duvakin. 2016. "Buried in a Head: African and Asian Parallels to AesopâĂŹs Fable." *Foklore*, 127: 91–102.

Boas, Franz. 1898. *Indianische Sagen von der Nordpazifischen Kuste Amerikas.* Berlin:Asher.

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

Boserup, Ester. 1970. *WomanâĂŹs Role in Economic Development.* London:George Allen and Unwin Ltd.

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

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

Chiovelli, Giorgio, Matteo Cervellati, and Elena Esposito. 2017. "Bite and Divide: Ancestral Exposure to Malaria and the Emergence and Persistence of Ethnic Diversity in Africa." LBS, mimeo.

Dell, Melissa, Nathan Lane, and Pablo Querubin. 2017. "The Historical State, Local Collective Action, and Economic Development in Vietnam." working paper, Harvard University, Department of Economics.

Diamond, Jared, and James A. Robinson. 2010. *Natural Experiments of History.* Cambridge, MA:Harvard University Press.

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

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

Evedy, Colin Mc, and Richard M. Jones. 1978. *Atlas of World Population History.* Harmondsworth and New York:Penguin.

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

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

Fernandez, Raquel. 2011. "Does Culture Matter?" In *Handbook of Social Economics.* Vol. 1A, , ed. Jess Benhabib, Matthew O. Jackson and Alberto Bisin, Chapter 11, 481–510. Elsevier, North-Holland.

Fortes, M., and E. Evans-Pritchard. 1940. *African Political Systems.* Oxford Press.

Foucault, Michel. 1995. *Discipline and Punish: The Birth of the Prison.* New York, NY:Vintage Books.

Franz Boas, Trans. D. Bertz. 2002. *Indian Myths and Legends from the North Pacific Coast of America.* Vancouver, Canada:Talonbooks.

Galor, Oded, and Omer Ozak. 2015. "Land Productivity and Economic Development: Caloric Suitability vs. Agricultural Suitability." mimeo, Brown University.

Galor, Oded, Omer Ozak, and Assaf Sarid. 2016. "Geographical Origins and Economic Consequences of Language Structures." Brown University Working Paper.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy. 2017. "Text as Data." NBER No. 23276.

Gentzkow, Matthew, Jesse Shapiro, and Matt Taddy. 2016. "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech." NBER No. 22423.

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

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

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

Green, Thomas A. 1997. *Folklore: An Encyclopedia of Beliefs, Customs, Tales, Music, and Art.* Vol. 1, Abc-clio.

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

Grosjean, Pauline. 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.

Gustav, Carl Jung. 1968. "The Archetypes and the Collective Unconscious." In *Collected Works of CG Jung.* Vol. 9. Princeton, NJ:Princeton University Press.

Harari, Yuval Noah. 2015. *Sapiens: A Brief History of Humankind.* New York, NY:HarperCollins.

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

Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey Sachs. 2004. "A Global Index of the Stability of Malaria Transmission." *American Journal of Tropical Medicine and Hygiene*, 70(5): 486–498.

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

Loughran, Tim, and Bill McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance*, 66(1): 35–65.

Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan Weigel. 2015. "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom." Econometrica, forthcoming.

Maddison, Angus. 2007. *Contours of the World Economy, 1-2030 AD; Essays in Macroeconomic History.* Oxford, United Kingdom:Oxford University Press.

Mayshar, Joram, Omer Moav, Zvika Neeman, and Luigi Pascali. 2015. "Cereals, Appropriability and Hierarchy." CEPR Discussion Paper 10742.

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

Michalopoulos, Stelios, Alireza Naghavi, and Giovanni Prarolo. 2017. "Trade and Geography in the Spread of Islam." The Economic Journal, forthcoming.

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

Michalopoulos, Stelios, and Elias Papaioannou. 2017*a*. "The Long Economic and Political Shadow of History." , ed. Stelios Michalopoulos and Elias Papaioannou, Chapter Series Introduction: Historical Legacies and Contemporary Development, viii–xxix. VOXEU.

Michalopoulos, Stelios, and Elias Papaioannou. 2017*b*. "Spatial Patterns of Development: A Meso Approach." forthcoming, Annual Review of Economics.

Michalopoulos, Stelios, Louis Putterman, and David N. Weil. 2016. "The Influence of Ancestral Lifeways on Individual Economic Outcomes in Sub-Saharan Africa." NBER WP 21907.

Murdock, George Peter. 1959. *Africa: Its Peoples and their Culture History.* McGraw-Hill.

Murdock, George Peter. 1967. *Ethnographic Atlas: A Summary.* PA:University of Pittsburgh Press.

Namenwirth, Zvi J., and Robert Philip Weber. 2016. *Dynamics of Culture.* Vol. Routledge Library Editions: Cultural Studies, Taylor and Francis.

Nisbett, Richard, and Dov Cohen. 1996. *Culture of Honor: The Psychology of Violence in the South.* Boulder, CO:Westview Press.

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

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

Osafo-Kwaako, Philip, and James A. Robinson. 2013. "Political Centralization in Pre-colonial Africa." *Journal of Comparative Economics*, 41(1): 534–564.

Richerson, Peter J., Monique Mulder, and Bryan Vila. 2001. *Principles of Human Ecology.* Chicago, IL:University of California - Davis.

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

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

Spolaore, Enico, and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, 51(2).

Stone, Phillip J., Dexter C. Dunphy, and Marshall S. Smith. 1966. "The General Inquirer: A computer Approach to Content Analysis."

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

Thompson, Stith. 1946. *The Folktale.* New York, NY:The Dryden Press.

Toelken, Barre. 1996. *Dynamics of Folklore.* University Press of Colorado.

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

Weber, Eugen. 1976. *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914.* Stanford University Press.

Wisniewski, Tomasz Piotr, and Brendan 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.

Witzel, Michael EJ. 2012. *The origins of the World's Mythologies.* Oxford University Press.

6 Appendix Figures and Tables

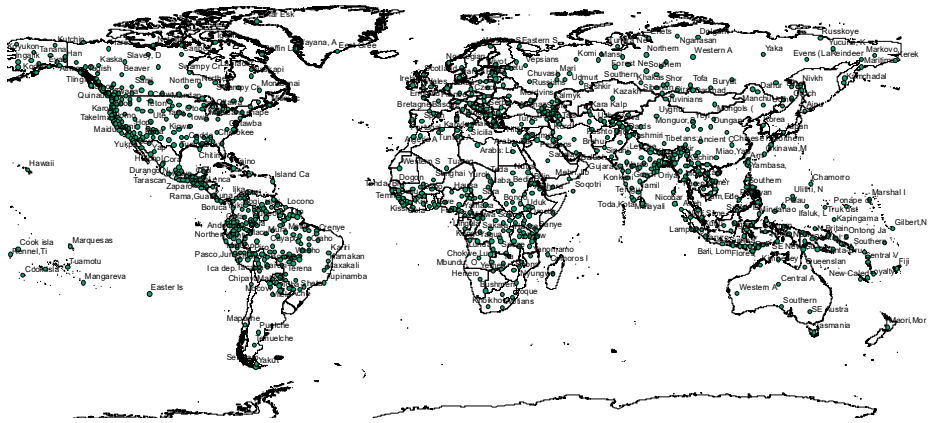


Figure 1: Societies in Berezkin

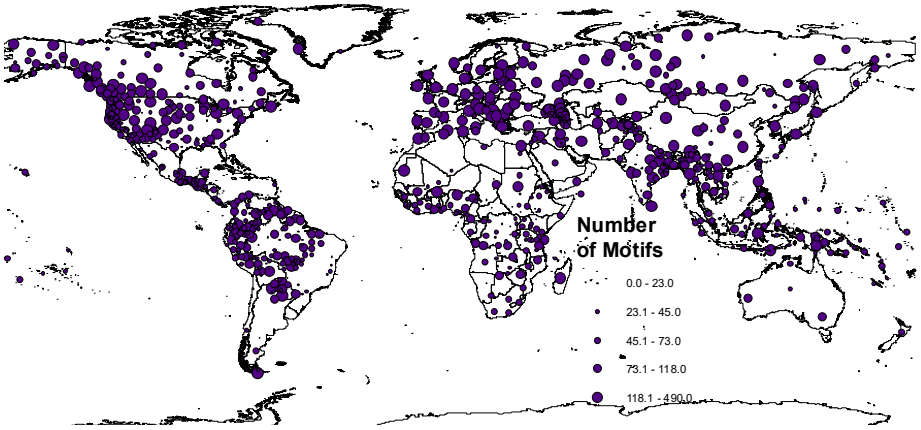


Figure 2: Number of Motifs Across Groups

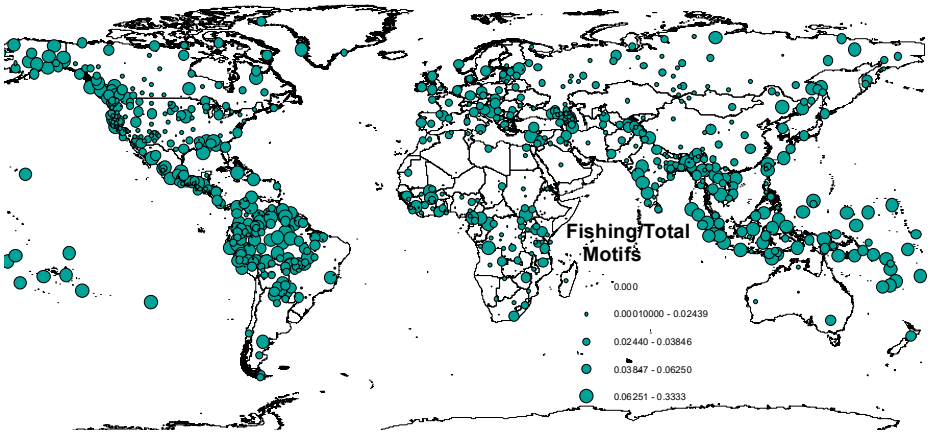


Figure 3: Fishing Motifs in the Oral Tradition

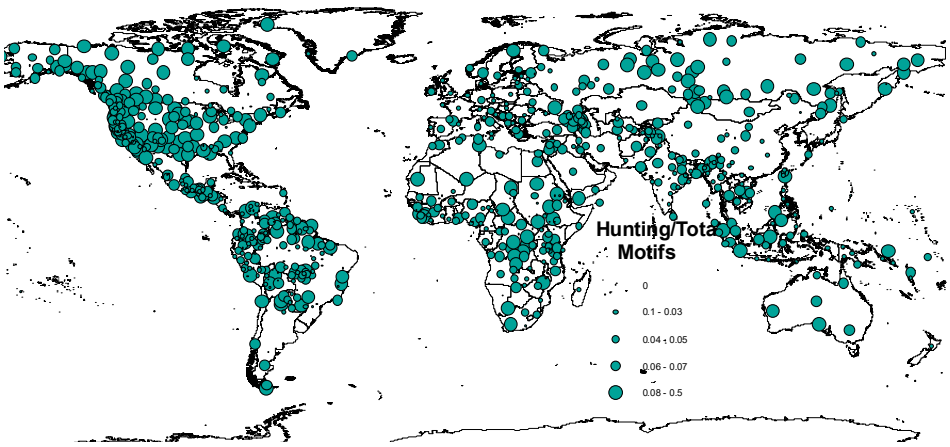


Figure 4: Hunting Motifs in the Oral Tradition

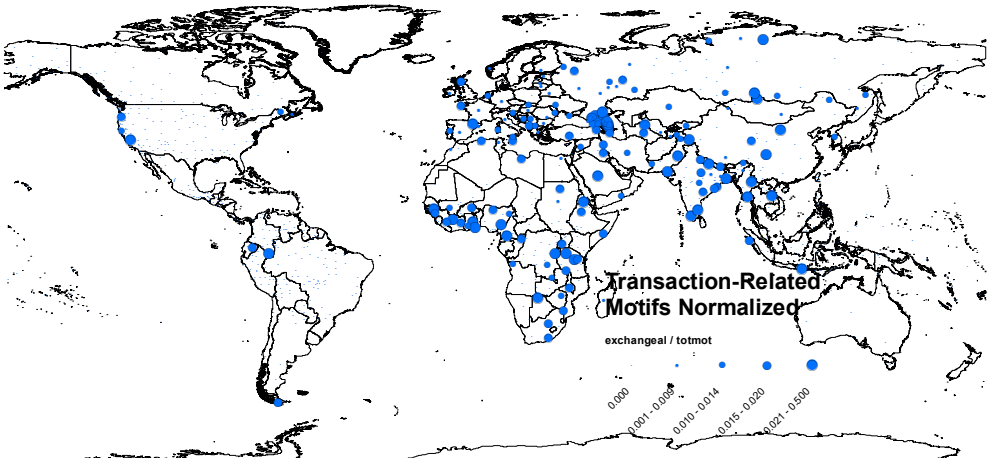


Figure 5: Exchange Economy in the Oral Tradition

Table 1 - Panel A: for Berezkin analysis

Variable	Mean	Std. Dev.	Min.	Max.	N
# of Motifs	79.791	75.538	1	516	936
# of Motifs on Earthquakes	0.103	0.34	0	3	936
# of Motifs on Rain & Thunder	3.022	2.932	0	17	936
# of Motifs on Mosquitoes	1.082	1.461	0	11	936
# of Motifs on Agriculture	1.53	2.27	0	16	936
# of Motifs on Fishing	2.595	3.112	0	17	936
Dist. to Earthquake Zones	576.92	716.882	0	4698.524	936
Dist. to the Coast	397.152	446.939	0.03	2293.282	936
Mean Malaria Intensity	2.789	5.686	0	31.317	935
Average Lightning Flash Density	10.182	10.561	0	71.543	933
Pre-1500 Agricultural Suitability	1361.404	1036.013	0	4888.25	923

Table 1: for Berezkin/EA analysis

Variable	Mean	Std. Dev.	Min.	Max.	N
# of Motifs on Farming	1.242	1.941	0	16	1233
# of Herding-Related Motifs	2.151	4.553	0	37	1233
# of Motifs on Hunting/Gathering/Fishing	7.084	6.772	0	41	1233
# of Motifs on Fishing	2.088	2.728	0	17	1233
# of Motifs on Hunting	3.334	3.551	0	21	1233
# of Hierarchy-Related Motifs	1.71	3.495	0	23	1233
# of Exchange-Related Motifs	1.564	3.234	0	26	1233
# of Motifs on 'Ought to'	1.906	2.77	0	19	1233
# of Motifs on Submission	5.725	6.544	0	50	1233
# of Motifs on Hostility	25.356	21.511	0	163	1233
# of Motifs	77.774	63.593	2	516	1233
Log(1st PC of Motifs Trickster Successful)	0	1.653	-3.07	4.814	1232
Log(1st PC of Motifs Trickster Unsuccessful)	0	1.519	-1.74	5.052	1232
Log(1st PC of Motifs Trickster Neutral)	0	1.618	-3.133	3.84	1232
Ln(1st PC of Trickster Motifs)	0	1.689	-3.954	4.322	1232
Degree of Political Complexity	1.889	1.036	1	5	1106
Animal Husbandry	1.549	1.801	0	9	1232
Agriculture	4.434	2.717	0	9	1232
Gathering	1.028	1.594	0	8	1232
Hunting	1.452	1.559	0	9	1232
Fishing	1.54	1.712	0	9	1232
Presence of Plow	0.119	0.324	0	1	1129
Role of Women has Declined	0.07	0.255	0	1	1233

Table 2 - Panel A: Breakdown of Sources used by Yuri Berezkin by Language

Afrikaans	65
Belarusian	12
Breton	6
Bosnian	19
Catalan/Valencian	108
Czech	9
Welsh	14
Danish	151
German	912
Greek (Modern)	3
English	5,650
Esperanto	40
Spanish	1,287
Estonian	108
Basque	15
Persian	2
Finnish	32
French	723
Western Frisian	40
Irish	9
Scottish Gaelic	6
Armenian	17
Hebrew (Modern)	1
Croatian	11
Hungarian	93
Armenian	1
Indonesian	18
Icelandic	5
Italian	43
Latin	106
Lithuanian	12
latvian	2
Nepali	97
Serbian	111
Norwegian	2
Polish	73
Portuguese	178
Quechua	4
Romanian	137
Russian	1,886
Sanskrit	14
Slovak	24
Slovenian	60
Albanian	11
Serbian	31
Swedish	40
Swahili	4
tagalog	4
Turkish	1
Vietnamese	31

Table 2 - Panel B: Top 10 Groups with the Largest Number of Motifs

Tradition	# of Motifs
Greek (modern)	371
Germans	372
Bashkir	379
Romanians	383
Kazakh	396
Lithuanians	409
Georgians	432
Bulgaria	484
Ukrainians	501
Russians	516

Table 2 - Panel C: Top Ten 10 Motifs with the Largest Number of Presence Across Groups

# of Traditions	ID	Description
277	a12	Some creature or creatures regularly (sunrise and sunset, summer and winter, lunar phases) or irregularly (solar and lunar eclipses, eschatological events) attack the luminaries or shade their light
279	b3b	Original earth was small and later increased in size or the fertile soil grew from a small amount of original substance
281	d4a	Fire is stolen from its original owner or brought back to the people from somebody who had stolen it before
290	k25	A man gets a woman connected with the upper world (bird-maiden, sky fairy, star-woman, etc.), she becomes his wife or (rare) adopted daughter
294	b3a	Water is the original element, the dry earth appears later
306	a5	The Moon is male, the Sun is also male or (much more rare) asexual
330	a3	The Moon is female or bisexual, the Sun is male
341	a32	A figure or an imprint of some being or object are seen in the Moon
343	k27n	Father or other kinsmen of hero's wife or bride try to kill or test him and/or suggest him difficult tasks
346	m29b	In episodes related to deception, absurd, obscene or anti-social behavior the protagonist is fox, jackal or coyote

Table 3 - Panel A: Folklore and the Physical Environment

	# of Motifs on Earthquakes		# of Motifs on Rain & Thunder		# of Motifs on Mosquitoes	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Dist. to Earthquake Zones)	-0.236***	-0.207***				
	(0.0370)	(0.0327)				
Ln(Average Lightning Flash Density)			0.228***	0.164***		
			(0.0394)	(0.0421)		
Ln(Mean Malaria Intensity)					0.190***	0.248*
					(0.0679)	(0.143)
Ln(# of Motifs)	0.490***	0.793***	0.791***	0.815***	0.973***	0.939***
	(0.0894)	(0.133)	(0.0411)	(0.0373)	(0.0574)	(0.0593)
Country FE	No	Yes	No	Yes	No	Yes
Continent FE	Yes	No	Yes	No	Yes	No
Log Likelihood	-291.4	-229.1	-1757.4	-1589.7	-1101.8	-972.3
Observations	936	936	933	933	935	935

Notes: The table reports Poisson estimates. Odd-numbered columns include continent-specific constants, even-numbered columns include country fixed effects (constants not reported). ***, **, * denote significance is 1%, 5%, and 10% level, respectively. Standard errors are clustered at the language level. See Data Appendix for variables definitions and Table 1 Panel A for summary statistics.

Table 3 - Panel B: Folklore, Subsistence and the Physical Environment

	# of Motifs on Agriculture		# of Motifs on Fishing	
	(1)	(2)	(3)	(4)
Ln(Pre-1500 Agricultural Suitability)	0.125***	0.111***		
	(0.0346)	(0.0239)		
Ln(Dist. to the Coast)			-0.104***	-0.0890***
			(0.0158)	(0.0196)
Ln(# of Motifs)	0.980***	1.033***	1.088***	1.114***
	(0.0520)	(0.0558)	(0.0507)	(0.0429)
Country FE	No	Yes	No	Yes
Continent FE	Yes	No	Yes	No
Log Likelihood	-1264.7	-1086.3	-1496.5	-1396.5
Observations	923	923	936	936

Notes: The table reports Poisson estimates. Odd-numbered columns include continent-specific constants, even-numbered columns include country fixed effects (constants not reported). ***, **, * denote significance is 1%, 5%, and 10% level, respectively. Standard errors are clustered at the language level. See Data Appendix for variables definitions and Table 1 Panel A for summary statistics.

Table 4 - Panel A: Folklore, Subsistence and the Ethnographic Record

	Agriculture		Animal Husbandry		Fishing		Hunting	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(1 + # of Herding-Related Motifs)	-0.216	-0.110	0.708***	0.463***	-0.157	-0.148	-0.155**	-0.0760
	(0.238)	(0.225)	(0.193)	(0.122)	(0.140)	(0.159)	(0.0719)	(0.0944)
Ln(1 + # of Farming-Related Motifs)	0.999***	0.570**	0.333***	0.287***	-0.355**	-0.251*	-0.605***	-0.395***
	(0.343)	(0.217)	(0.103)	(0.0905)	(0.160)	(0.131)	(0.133)	(0.138)
Ln(1 + # of HGF-Related Motifs)	-0.908***	-0.464***	0.240	-0.0261				
	(0.280)	(0.171)	(0.169)	(0.0757)				
Ln(1 + # of Fishing-Related Motifs)					0.695***	0.599***		
					(0.213)	(0.200)		
Ln(1 + # of Hunting-Related Motifs)							0.505***	0.357***
							(0.116)	(0.112)
Ln(# of Motifs)	0.364	0.379**	-0.546**	-0.165*	-0.204	-0.321*	0.0648	-0.0494
	(0.271)	(0.182)	(0.266)	(0.0947)	(0.148)	(0.163)	(0.0846)	(0.0823)
Country FE	No	Yes	No	Yes	No	Yes	No	Yes
Continent FE	Yes	No	Yes	No	Yes	No	Yes	No
Adjusted R^2	0.357	0.537	0.387	0.572	0.268	0.352	0.456	0.521
Observations	1232	1232	1232	1232	1232	1232	1232	1232

Notes: The table reports OLS estimates. Odd-numbered columns include continent-specific constants, even-numbered columns include country fixed effects (constants not reported). ***, **, * denote significance is 1%, 5%, and 10% level, respectively. Standard errors are clustered at the language level. See Data Appendix for variables definitions and Table 1 Panel B for summary statistics.

Table 4 - Panel B: Folklore, Institutions and Exchange

	Degree of Political Complexity		Ln(1 + # of Exchange-Related Motifs)		
	(1)	(2)	(3)	(4)	(5)
Ln(1 + # of Hierarchy-Related Motifs)	0.562***	0.442***			
	(0.128)	(0.104)			
Ln(1+Distance to the Trade Routes, 600 AD)			-0.149***	-0.142***	-0.115***
			(0.0298)	(0.0426)	(0.0322)
Ln(1+Distance to the Trade Routes, 1700 AD)					-0.0452
					(0.0404)
Ln(# of Motifs)	-0.150*	-0.0632	0.595***	0.518***	0.517***
	(0.0753)	(0.0493)	(0.0668)	(0.0717)	(0.0712)
Country FE	No	Yes	No	Yes	Yes
Continent FE	Yes	No	Yes	No	No
Adjusted R^2	0.306	0.414	0.600	0.688	0.688
Observations	1106	1106	774	774	774

Notes: The table reports OLS estimates. Column 1 & 3 include continent-specific constants, column 2, 4 & 5 include country fixed effects (constants not reported). ***, **, * denote significance is 1%, 5%, and 10% level, respectively. Standard errors are clustered at the language level. See Data Appendix for variables definitions and Table 1 Panel B for summary statistics.

Table 5 - Panel A: Folklore and Long-Standing Conjectures in Anthropology

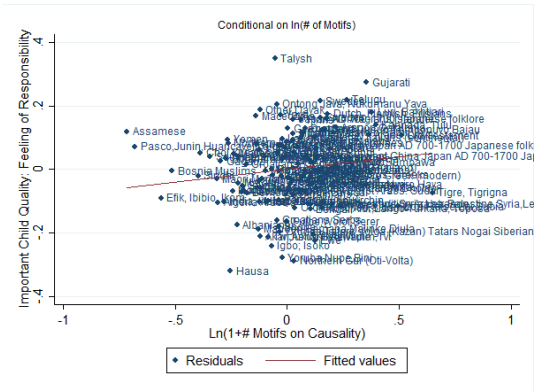
	Ln(1 + # of Exchange-Related Motifs)			Role of Women Has Declined		Ln(1+# of Motifs on Hostility)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Degree of Political Complexity	0.148***	0.0998***	0.0878***				
	(0.0373)	(0.0313)	(0.0296)				
Agriculture			0.00107	-0.00552	0.000229	-0.00806**	0.00185
			(0.0127)	(0.00427)	(0.00622)	(0.00359)	(0.00461)
Animal Husbandry			0.0630***			0.0262***	0.0173**
			(0.0179)			(0.00396)	(0.00768)
Presence of Plough				0.118**	0.0929		
				(0.0529)	(0.0745)		
Ln(# of Motifs)	0.458***	0.395***	0.388***	0.0192	0.0286*	0.972***	0.943***
	(0.0658)	(0.0560)	(0.0544)	(0.0133)	(0.0161)	(0.0167)	(0.0202)
Country FE	No	Yes	Yes	No	Yes	No	Yes
Continent FE	Yes	No	No	Yes	No	Yes	No
Adjusted R^2	0.540	0.647	0.657	0.0603	0.253	0.922	0.935
Observations	1106	1106	1106	1129	1129	1232	1232

Notes: The table reports OLS estimates. Column 1, 4 & 6 include continent-specific constants, column 2, 3, 5 & 7 include country fixed effects (constants not reported). ***, **, * denote significance is 1%, 5%, and 10% level, respectively. Standard errors are clustered at the language level. See Data Appendix for variables definitions and Table 1 Panel B for summary statistics.

Table 5 - Panel B: Rule-Following and the Historical State: Evidence from Folklore

	Degree of Political Complexity						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ln(1+ # of Motifs on "Ought to")	0.336**	0.269***					
	(0.134)	(0.0887)					
Ln(1+ # of Motifs on Submission)			0.312***	0.191**			
			(0.101)	(0.0806)			
Ln(1st PC of Trickster Motifs)					0.103		
					(0.0874)		
Log(1st PC of Motifs Trickster Successful)						0.0703	0.0558
						(0.0587)	(0.0676)
Log(1st PC of Motifs Trickster Unsuccessful)						0.213***	0.166**
						(0.0576)	(0.0649)
Log(1st PC of Motifs Trickster Neutral)						-0.115**	-0.108**
						(0.0540)	(0.0459)
Ln(# of Motifs)	-0.0825	-0.0500	-0.151	-0.0567	-0.0863	-0.122	-0.0435
	(0.0885)	(0.0675)	(0.0987)	(0.0756)	(0.170)	(0.145)	(0.100)
Country FE	No	Yes	No	Yes	No	No	Yes
Continent FE	Yes	No	Yes	No	Yes	Yes	No
Adjusted R^2	0.249	0.394	0.239	0.385	0.227	0.275	0.404
Observations	1106	1106	1106	1106	1106	1106	1106

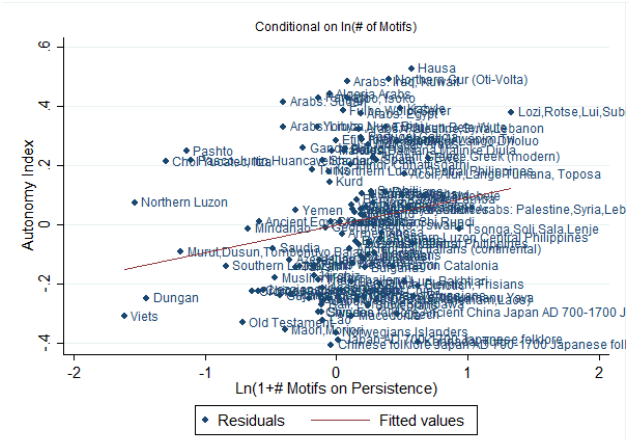
Notes: The table reports OLS estimates. Column 1, 3, 5 & 6 include continent-specific constants, column 2, 4 & 7 include country fixed effects (constants not reported). ***, **, * denote significance is 1%, 5%, and 10% level, respectively. Standard errors are clustered at the language level. See Data Appendix for variables definitions and Table 1 Panel B for summary statistics.



(a): Causality and Responsibility



(b): Submission and Autonomy



(c): Perseverance and Autonomy



(d): Work, good? Work, bad?

Figure 1: Folklore and Contemporary Attitudes