Values of China: Toward a Cultural Map*

Melanie Meng Xue†

LSE

January 2026

Abstract

This paper constructs a county-level cultural map of China using nearly 20,000 folktales collected nationwide in the 1980s. I develop a synthetic-archetype approach that combines narrative prompts with sentence-level BERT embeddings to measure the salience of values and informal social-order narratives—such as honesty, trust, effort, patience, kinship, religion, reciprocity, and beliefs about political authority—embedded in folktales. I assess reliability using stability across independent archetype draws and convergence with an independent LLM classifier, and I use style controls to separate thematic content from narrative form. After aggregating the measures to the county level, I document systematic historical and geographic correlates, linking cultural narratives to ecological conditions, religious competition, and long-run state presence. The resulting indices predict related contemporary attitudes in the Chinese General Social Survey, including trust, prosocial behavior, beliefs about effort, intertemporal orientation, religious affiliation, kin obligations, gift pressures, and trust in state authority. In models predicting county GDP per capita with rich historical, geographic, and contemporary controls and Lasso selection, patience, karma, and honesty–trust enter positively, while clan narratives, ancestor worship, and benevolent-emperor themes enter negatively.

Keywords: Culture, History, Growth, China
JEL Codes: N35 O40 Z1

*I thank Andrew Bernard, Loren Brandt, Martin Fiszbein, Guido Friebel, Paola Giuliano, Stephan Heblich, Laurence Iannaccone, Sara Lowes, Andrei Markevich, Joel Mokyr, Gerard Padro, Moses Shayo, Uwe Sunde, Guido Tabellini, Hans-Joachim Voth, Tianyi Wang, David Yang, Fabrizio Zilibotti and Xiaodong Zhu for valuable comments. I also thank seminar audiences at Aix-Marseille School of Economics, Fudan University, Helsinki GSE, PKU INSE, University of Toronto and Utrecht University, as well as audiences at the BSE Summer Forum 2025, CESifo Venice Summer Institute 2025, Harvard Workshop on Historical Kinship, Psychology, and Economic Prosperity (2023), HKU Conference on Asia and the Global Economy (2024), KCL Workshop in Comparative Politics and Political Economy (2024), and the Oxford–Tsinghua Conference on Chinese Economic History from a Comparative Perspective (2024) for helpful feedback. I am especially grateful to Runping Zhao and Congshan Shen for excellent research assistance. I also thank Jinghan Chang, Qiujin Chen, Shanshan Li, and Dong Xue, and the teams they respectively led, for their help during the early experimentation stage of the project, when the work relied mainly on hand-coding and traditional NLP methods prior to the availability of large language models.

†Assistant Professor at London School of Economics and Political Science (LSE) and Research Affiliate

I Introduction

Culture significantly shapes decision-making processes and societal structures, influencing economic performance, governance, and institutional development (Boyd and Richerson, 1985; Putnam, 1993; Boyd and Richerson, 2005; Nunn, 2012). Research has shown that cultural values—such as trust, social capital, individualism, and reciprocal norms—are particularly important in promoting economic growth and political stability, as demonstrated in historical contexts like the Industrial Revolution and in modern development (Tabellini, 2008; Gorodnichenko and Roland, 2017; Besley, 2020).

This perspective aligns with prior research emphasizing how societal beliefs and narratives shape long-run economic choices, social norms, and institutional structures (Bénabou and Tirole, 2006; Fernández, 2011). In particular, internalized cultural values—whether about honesty, effort, or reciprocity—can persist through intergenerational transmission and influence policy preferences, labor-market participation, and collective behaviors.

Comparative analyses of cultural and institutional evolution in Europe and China have also provided critical perspectives on divergent developmental pathways (Greif and Tabellini, 2017; Mokyr and Tabellini, 2024). For instance, Greif and Tabellini (2017) analyze how China's clan-based social structures fostered collective decision-making and hierarchical organization, contrasting with Europe's individualistic frameworks that supported the emergence of democratic governance and market-based economies. Similarly, Mokyr and Tabellini (2024) analyze how differences in social organization shaped technological innovation and economic progress, highlighting the role of culture in driving development.

Despite substantial advancements, systematically measuring cultural traits across time and space remains a significant challenge. This is especially true in the context of China, where the vast geographic and temporal scale, combined with limited data on localized cultural practices, creates a gap in understanding how these traits are distributed and how they relate to modern outcomes. Although a nascent literature has started to examine cultural narratives as a means of capturing enduring beliefs and practices (Michalopoulos and Xue, 2021), efforts to systematically quantify these traits at the local level, particularly within China, remain scarce.

China offers favorable conditions for studying historical cultural values, including extensive written records, diverse local traditions, and relatively recent urbanization. However, much of this potential remains unrealized due to historical barriers to anthropological research during the socialist period and the lack of systematic efforts to measure cultural traits

at the Centre for Economic Policy Research (CEPR). Email: m.m.xue@lse.ac.uk

across regions. Modern datasets, such as the CGSS, provide only limited geographic coverage and offer an incomplete picture of the country's historical cultural traits and their regional variation. As a result, much of China's historical and regional variation in cultural traits remains underexamined.

These gaps are particularly striking given the historical continuity of cultural traits in China. Despite rapid modernization, much of the country's traditional social fabric remains intact due to relatively recent urbanization. Scholars such as Perkins et al. (1975) and Richardson (1999) have highlighted the importance of these enduring cultural traits in shaping local governance, economic behavior, and social interactions. Historical narratives, captured in folktales and other cultural artifacts, provide a crucial lens for understanding these continuities and their modern relevance.

To address these challenges, this study analyzes a dataset of nearly 20,000 folktales collected in the 1980s. These narratives were digitized and standardized, creating a structured dataset suitable for systematic analysis. Using this dataset, the study employs a methodological framework to quantify cultural themes embedded in the folktales. Synthetic archetypes representing key cultural topics—such as honesty, trust, charity, Confucianism, and gift exchange—serve as benchmarks for evaluating thematic similarity.

The thematic analysis leverages BERT embeddings to calculate similarity scores between the collected folktales and these synthetic benchmarks. Additionally, a large language model (LLM) classifies the thematic content of each tale into predefined cultural categories, offering a scalable and efficient alternative method for analyzing and categorizing thematic content. Together, these approaches provide complementary insights into how historical narratives reflect cultural values across regions.

By aggregating the similarity scores at the county level, the study links historical cultural traits to modern attitudes and behaviors using datasets such as the CGSS. Furthermore, it examines the relationship between these cultural traits and regional economic performance, specifically exploring whether historical narratives align with contemporary differences in GDP per capita.

This study makes several contributions. First, it provides a systematic cultural map of China, offering quantified measures of cultural traits across regions using historical narratives. By aggregating cultural themes derived from nearly 20,000 folktales, this research creates a geographically and thematically detailed representation of China's cultural characteristics. This cultural map fills a significant gap in the literature, where within-country analyses of cultural variation often lack systematic and consistent measures across multiple dimensions

of culture. The framework established here enables future studies to explore local cultural traits and their relationships to economic and political phenomena.

Second, the research introduces historical cultural traits as an additional explanatory factor for regional economic disparities in China. A large literature on China's growth and uneven regional development emphasizes institutional reforms, structural change, and the spatially uneven evolution of policy and factor markets (Lin, 1995; Brandt and Rawski, 2008; Song, Storesletten, and Zilibotti, 2011; Zhu, 2012; Kanbur and Zhang, 2005; Wang and Yang, 2025). Complementing these accounts, this study shows that historically rooted cultural traits inferred from folktales, such as honesty and effort, are systematically associated with contemporary outcomes that matter for development, including trust in state institutions and county-level GDP per capita. Evidence on cross-society preference heterogeneity (Falk et al., 2018) supports the idea that durable differences in trust, patience, and reciprocity can be economically consequential. By incorporating cultural factors into the analysis, the study expands the set of candidate mechanisms behind China's uneven growth and helps link place-based cultural heterogeneity to modern economic performance.

Third, the methodological framework used in this research contributes to the growing field of cultural quantification. While survey-based measures of cultural traits—such as trust or reciprocity—are well-documented, this study demonstrates how qualitative data from historical narratives can be systematically analyzed to quantify cultural themes. It is the first to apply generative AI to the analysis of folktales, using synthetic stories to construct reference archetypes that encode culturally positive portrayals of specific traits. By combining machine learning methods—such as BERT embeddings and LLM-based classification—with large-scale historical datasets, the research offers a replicable approach for analyzing qualitative cultural data. This methodology can be applied to other regions, enabling comparative studies of cultural persistence and its economic implications.

Fourth, this study advances research on traditional religions and superstitions by analyzing how these beliefs were embedded in historical narratives and their relevance to modern economic and social outcomes. While much existing literature focuses on monotheistic traditions, this study highlights the role of indigenous Chinese belief systems—including ancestor worship, Buddhist influences, and supernatural themes in folklore—in shaping cultural attitudes. In the China context, much of the economics literature has emphasized Confucianism as a state-linked ideological tradition with implications for values and institutional development.[1] By contrast, comparatively less is known about how vernacular religious and

[1] Jia and Kung (2025) survey the culture and institutions of Confucianism and synthesize research on its relationship to clan culture, state institutions, and a wide range of societal values.

supernatural beliefs were expressed in historical narratives and how these beliefs relate to contemporary outcomes. Prior research has shown how religious beliefs influence economic incentives and social structures (Iannaccone, 1998; Barro and McCleary, 2003; Becker, Rubin, and Woessmann, 2024) and how supernatural beliefs can function as informal institutions that regulate behavior and cooperation (Nunn and Sierra, 2017). Von Glahn (1991) and Von Glahn (2004) examine Wutong, a morally ambivalent wealth god, who granted fortune but often at ethically troubling costs, reflecting anxieties over commerce and economic risk in Song China. Additionally, Nye and Xue (2014) demonstrate how cultural conceptions of auspiciousness—such as the Year of the Dragon—can alter demographic patterns. By quantifying these cultural traits, this study provides new evidence on their historical persistence and their association with contemporary beliefs and regional economic development.

Finally, the study contributes to broader discussions on how historical cultural traits shape long-term development. Comparative research has examined how cultural and institutional evolution diverged between Europe and China (Greif and Tabellini, 2017; Mokyr and Tabellini, 2024), shaping distinct developmental trajectories. This study shifts the focus to China's regional variation, showing that cultural traits documented in historical narratives remain relevant for understanding contemporary economic differences. By providing systematic evidence of regional cultural variation within a single country, it complements existing cross-country analyses and highlights the enduring influence of historical beliefs and values.

The remainder of this paper is organized as follows. Section II describes the data and methodologies employed in this study. Section III analyzes the values embedded in historical folktales and their connections to contemporary societal attitudes and behaviors. Section IV explores the relationship between these cultural traits and economic growth. Section V discusses other historical sources, such as local gazetteers and poetry, which capture cultural variation over time and could be explored in future research. Section VI concludes by summarizing the findings, discussing their contributions to the literature, and outlining directions for future research. Additional methodological details and supporting materials are included in the Appendix.

II Data and Methodology

II.A Why Folktales?

This study analyzes regional variation in cultural traits in China using historical narratives, specifically folktales. A long tradition in folklore and anthropology emphasizes oral narratives as repositories of social norms, moral expectations, and shared beliefs that struc-

ture everyday behavior (Bascom, 1965; Dundes, 1980). Because folktales are transmitted across generations through oral storytelling, they tend to preserve cultural content formed prior to major institutional or ideological shifts, offering access to older layers of moral reasoning and social values that are difficult to recover from modern surveys (Lévi-Strauss, 1962; Malinowski, 1926).

A defining feature of folktales is that moral principles are embedded in narrative structure rather than stated explicitly. Ethical rules are conveyed through plot and causal resolution: virtuous behavior is rewarded, wrongdoing is punished, and supernatural forces often serve as the mechanism through which moral order is enforced. These narrative devices make norms salient and memorable, facilitating their transmission and internalization even in settings with limited formal enforcement.

Unlike official written sources, which often reflect elite perspectives or state priorities, folktales emerge from everyday social contexts. They capture locally grounded understandings of trust, effort, reciprocity, authority, and obligation as negotiated in daily life (Finnegan, 1992). Their local embeddedness, combined with systematic regional collection, makes folktales particularly well suited for studying persistent cultural variation across space. In China, folktale collections are organized by province and county, allowing for fine-grained geographic comparisons.

Finally, folktales address an important data limitation in the Chinese context. Large-scale surveys of beliefs and values are largely unavailable prior to the reform era, and twentieth-century anthropological fieldwork was uneven across regions (Eberhard, 1952). By quantifying thematic content in historical folktales, this study constructs measures of cultural traits that predate modern economic development and institutional change, while remaining comparable across regions. This approach complements existing work that emphasizes institutions, geography, and policy in explaining China's regional disparities.

II.B The Folktale Collection Project

In the 1980s, China undertook a large-scale effort to document folktales that had been transmitted orally for generations. This work formed part of the "Three Collections of Chinese Folk Literature" initiative launched in 1984. The project involved extensive fieldwork across provinces and counties and resulted in a large archive of narratives that were widely viewed at the time as being at risk of disappearance amid rapid social and economic change.

The timing of the project was favorable for capturing orally transmitted material. China remained predominantly rural, with GDP per capita below $350, and many communities continued to rely on locally embedded practices and belief systems. Limited urbanization and

restricted access to mass media meant that oral traditions often persisted in relatively stable form. As a result, a substantial share of the recorded narratives plausibly reflect cultural material formed prior to the major institutional, demographic, and ideological transformations of the reform era.

The collection also coincided with the post–Cultural Revolution reconstruction of folklore studies as an academic field. After a prolonged disruption of research, scholars emphasized systematic documentation, large-scale fieldwork, and archival preservation as central tasks of the discipline. In this context, folktale collection was understood primarily as an exercise in documentation rather than interpretation, with particular attention paid to recording region-specific traditions before ongoing social and economic change altered the conditions of oral transmission.

Contemporary project documentation instructed collectors to record tales as told, avoid modernization of language or content, and preserve local variants rather than harmonizing narratives across places. A central emphasis of folklore research in this period was attention to local variation, with the explicit aim of documenting geographically specific traditions rather than producing standardized accounts.

The archive was assembled through collaboration among folklore societies, local cultural organizations, and community participants. In practice, the project relied on a division of labor between storytellers, who recounted local narratives, and collectors, who transcribed and organized them. Collectors came from diverse backgrounds, including teachers, local cultural workers, and academically trained researchers, contributing to heterogeneity in recording practices across localities. The existence of a shared written language across regions facilitated transcription despite substantial dialect variation.

Despite its scale and historical importance, the resulting archive has seen limited use in quantitative social science. To date, there has been no systematic effort to examine these materials at a national scale or to link their thematic content to contemporary regional outcomes. This study draws on the breadth and geographic structure of the collection to construct comparable measures of historical cultural traits.

Tales Recorded The folktale collection project spanned all provinces in China, documenting a total of 18,228 tales. Each province contributed a dedicated volume, forming a geographically organized archive. The number of tales contributed by each province ranges from 335 to 1,128.

Most of these tales can be traced to specific counties. Out of the 18,228 total tales, 17,070 are linked to particular counties, while some tales from certain ethnic minority groups

cannot be precisely mapped. Since these regions are not included in the current sample, this distinction does not affect the analysis.

The folktales cover diverse themes, including myths, legends, and local stories. Examples range from *Pangu Separates Heaven* and *Earth to The Butterfly Lovers*. On average, each folktale comprises about 981 words in length.

Storytellers and Collectors The compilation of folktales drew on local storytellers, many of whom were peasants in rural China. The average age of storytellers is 59.5 years.[2] In the storyteller metadata, 3,035 contributors are coded as illiterate and 4,097 report at most elementary schooling (excluding illiteracy), which is consistent with very low schooling among older cohorts. Peasant storytellers contributed 6,978 tales. These backgrounds suggest that the recorded narratives are more likely to reflect oral transmission than formal schooling or print-based literary exposure, which helps motivate their use as proxies for locally embedded cultural traits (Ong, 1982; Vansina, 1985; Lord, 1960). Figure A.2 reports the distribution of storyteller occupation and education.

Collectors who recorded the tales tend to have more formal schooling and are younger on average, with a mean age of 36.9 years. Among collectors, 186 report elementary schooling or below while remaining literate, 5,942 report secondary schooling (junior high, senior high, or vocational), and 4,098 report college or above. This profile is unsurprising given that the tasks of locating, transcribing, and organizing folktales typically required literacy and record-keeping skills that were more common among teachers and other educated intermediaries.

Ethnic Tales While the full folktale corpus includes narratives from China's 56 officially recognized ethnic groups, this study focuses exclusively on Han Chinese traditions. Minority narratives are excluded due to both conceptual and empirical limitations. Many minority tales lack geographic identifiers, preventing linkage to local outcomes, and available survey data—such as the CGSS—offer limited coverage of non-Han populations.

The exclusion allows for a more coherent analysis of Han cultural patterns using the proposed methodology. Descriptive statistics for the excluded ethnic tales are provided in Appendix C, which also discusses the potential for future research building on these materials.

[2] Authors' calculations from the 1% microdata sample of the 1982 China Population Census made available through IPUMS International. In these data, more than 70% of individuals aged 59 are coded as illiterate. See the IPUMS International documentation for the China 1982 sample design and variable definitions (IPUMS International, 2025).

II.C Quantifying Cultural Themes in Folktales

The corpus consists of long, coherent narratives. Scanned texts are OCRed and lightly cleaned to remove headers and standardize punctuation. Chinese folklore lacks a fine-grained index comparable to ATU or the Berezkin Catalogue (Uther, 2004; Berezkin, 2015), whose coverage within China is too coarse for regional analysis.[3] Because folktales convey values through roles, intentions, and causal resolution rather than keywords, bag-of-words and LDA perform poorly on single-theme narratives. Using narrative representations aligns with the object of interest and with evidence that narratives carry economically relevant content (Shiller, 2017).

For each theme, I generate several Synthetic Archetypes with instruction-following prompts that fix the target concept and match the register of Chinese folktales. I pre-compute embeddings for these archetypes and, for each theme, measure the semantic proximity of a tale to the theme by embedding the tale with a Chinese Sentence-BERT model and averaging cosine similarities across the theme's archetypes. Narrative SAs are construct-valid for narrative data and match encoder training on natural texts (Reimers and Gurevych, 2019). The design follows text-as-data practice—multiple exemplars, external validation, and robustness—and treats narratives as economically meaningful units (Gentzkow, Kelly, and Taddy, 2019; Shiller, 2017).

Themes are of two types. Virtues such as honesty, trust, charity, effort, and patience are prompted so that the protagonist exemplifies the value and is rewarded, hence higher similarity indicates both presence and positive affirmation. Institutions and practices such as clan organization, ancestor worship, Buddhism, reciprocal gift exchange, and the benevolent-emperor ideal are prompted for presence without evaluation, hence higher similarity indicates salience rather than appraisal. Tale-level scores are averaged to the county to form regional indices for each theme.

III Cultural Traits: Measures and Validation

Folktales collected during the 1980s—many of which originated centuries earlier—offer a window into the cultural values and beliefs of historical Chinese society. Because these narratives were transmitted, adapted, and retold across generations, they encode shared cultural orientations rather than idiosyncratic views of individual authors. The tales span a wide range of themes, from moral values that appear across many societies, such as honesty,

[3]The Berezkin Catalogue attributes motifs to broad regions in China and covers only a small set of Han traditions, which limits within-country inference.

prosociality, effort, and patience, to beliefs and practices that have been especially salient in China's historical and philosophical traditions, including Buddhism, ancestor worship, clan organization, and ritual gift exchange.

This section introduces three families of cultural measures constructed from folktales. The full set of prompts and county-level maps for all measures are reported in Appendix 3,

III.A Three Families of Cultural Measures

All measures rely on the same semantic-alignment pipeline described in Section II. In brief, I generate Synthetic Archetype (SA) stories using fixed prompts, embed both folktales and SAs with a Chinese Sentence-BERT model, compute cosine similarities, average across independently generated SAs, and aggregate to the county level. The families differ in the conceptual domain they target and in how prompts are framed.

Values and behavioral preferences. The first family captures foundational values and behavioral preferences that underpin social interaction and economic life. I focus on five dimensions emphasized in recent work on culture, growth, and institutional performance (e.g., Falk et al., 2018; Schulz et al., 2019; Inglehart, 2018): honesty, trust, altruism, effort, and patience.

Honesty and trust facilitate exchange by reducing enforcement costs. Inherited trust predicts higher growth and investment (Zak and Knack, 2001; Algan and Cahuc, 2010), and field evidence documents honest behavior even when dishonesty would be privately profitable (Cohn et al., 2019). Altruism—voluntary actions that benefit others at a personal cost—supports mutual aid and reciprocity; both "warm-glow" preferences (Andreoni, 1990) and reputational mechanisms (Bénabou and Tirole, 2006) help sustain prosocial contributions. Effort, understood as industriousness and perseverance, is closely linked to human-capital accumulation and labor-market success (Bowles, Gintis, and Osborne, 2001; Heckman and Rubinstein, 2001). Patience shapes intertemporal choice in growth theory (Galor and Özak, 2016) and is empirically associated with higher education, income, and savings at the societal level (Golsteyn, Grönqvist, and Lindahl, 2014; Falk et al., 2018).

For this family, prompts require that the value be exemplified and positively rewarded within the narrative. As a result, higher scores reflect both the presence of the value and its affirmative moral framing rather than mere topical salience. The exact prompts and county-level maps for honesty, altruism, effort, and patience are reported in Appendix 3.

Beliefs and practices specific to China. The second family captures traditional beliefs and practices that have played a central role in China's social order. I focus on Buddhism,

ancestor worship, clan organization, and ritual gift exchange. While none of these practices is unique to China, each has been especially prominent in its historical experience and continues to shape moral expectations, coordination, and authority.

Buddhist temples and lay networks diffuse a karmic ethic linking present intent to future consequences, fostering compassion, honesty, and self-discipline where secular enforcement is weak; at the same time, karmic worldviews can rationalize misfortune as deserved fate, potentially dampening risk-taking and experimentation (Keown, 1992; Gombrich, 2006; White and Willard, 2025). Ritual veneration of ancestors strengthens lineage identity and moral accountability, but its backward-looking orientation may slow organizational change and technological adoption; regions where ancestral rites remained especially strong appear culturally more resistant to modernization (Greif, Mokyr, and Tabellini, 2025). Clan-based governance provides insurance and in-group sanctioning while narrowing the radius of trust; cultural-evolution and economic-history research links dense kin networks to weaker impersonal cooperation and less inclusive institutions (Henrich, 2020; Greif and Tabellini, 2017). Gift exchange is a canonical form of positive reciprocity across societies (Falk et al., 2018); in China, reciprocal gifting also signals status and sustains social networks, yet the same norm underlies *guanxi* and can blur into favoritism or petty corruption (Mauss, 1954; Yang, 1994; Wedeman, 2012).

For beliefs and practices, prompts specify narrative presence without moral evaluation. Accordingly, higher scores capture *salience* rather than endorsement. Prompts and county-level maps for Buddhism, ancestor worship, clan organization, and gift exchange are reported in Appendix 3.

Legitimacy of central authority. The third family measures beliefs about the legitimacy of centralized state authority. Political economists and anthropologists emphasize that when citizens regard state institutions as rightful and trustworthy, compliance becomes quasi-voluntary, enforcement costs fall, and public-goods provision improves (Lipset, 1959; Levi, 1988; Besley and Dray, 2024). In China's narrative tradition, such beliefs are often expressed through tales of the benevolent emperor, in which a just ruler ensures prosperity and social order.

I therefore construct a county-level salience measure of the benevolent-emperor motif. Higher values indicate that this motif is prominent in local narrative traditions rather than that it is normatively endorsed by all storytellers. The prompt and county-level map are shown in Figure A.6; full documentation will be included in Appendix 3.

III.B Worked Example: Trust

To make the construction concrete, I illustrate the procedure using *trust*. I generate five Synthetic Archetype stories from a fixed prompt template and compute cosine similarities between each folktale and each SA using a Chinese Sentence-BERT model. Similarities are averaged across the five SAs and then aggregated to the county level.

Trust prompt—The story should mention interpersonal trust or suspicion, whether trust is being built, maintained, or lost in complex situations. Specifically, the narrative should illustrate how characters establish or preserve trust to foster cooperation and social cohesion.

Figure I presents the resulting county-level trust measure.

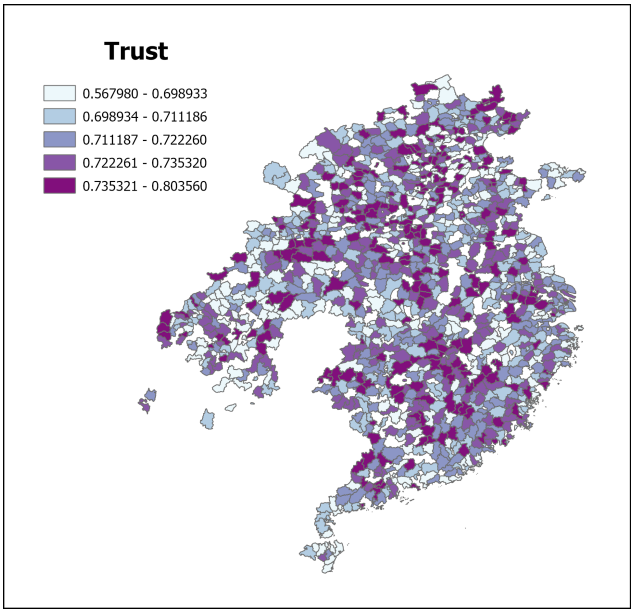


Figure I: County-Level Trust Narratives from Folktales

Note: The figure shows the county-level SA–BERT similarity measure for trust, constructed as the average semantic alignment between local folktales and independently generated trust archetypes. Darker shading indicates greater salience of trust-related narratives.

III.C Robustness and Validation

I assess the reliability and interpretability of the folktale-based measures using two complementary exercises. First, I examine stability across independent Semantic Archetype (SA) replications generated from the same prompt. Second, I compare SA–BERT similarity scores to an independent LLM-based classifier applied to the same folktales. All analyses in this subsection restrict attention to Han Chinese tales; ethnic-minority tales are excluded. The resulting sample consists of 14,535 tales.

Stability across SA replications. To assess internal reliability, I compute Pearson correlations among five independent SA–BERT replications generated from the same prompt. Table I reports these correlations for *trust*. Panel A uses raw similarities. Panel B reports correlations after residualizing each replication on the Neutral-SA similarity only, which partials out generic folktale register common across narratives. High correlations in both panels indicate that the extracted signal is stable to stochastic variation in SA generation and is not driven by narrative style.

Across values, the five SA replications are highly concordant: the mean pairwise correlation across replications is approximately 0.87, and Cronbach's alpha for the simple average is 0.97. Results are unchanged if I use the first principal component across replications instead of the mean.

Table I: Stability Across Five SA Replications for Trust

Panel A. Raw					
	Version 1	Version 2	Version 3	Version 4	Version 5
Version 1	1.0000				
Version 2	0.8567***	1.0000			
Version 3	0.6226***	0.4261***	1.0000		
Version 4	0.7421***	0.6880***	0.5145***	1.0000	
Version 5	0.8875***	0.8030***	0.6454***	0.5985***	1.0000
Panel B. Residualized on Neutral SA					
	Version 1	Version 2	Version 3	Version 4	Version 5
Version 1	1.0000				
Version 2	0.8110***	1.0000			
Version 3	0.4844***	0.2457***	1.0000		
Version 4	0.6351***	0.5806***	0.3265***	1.0000	
Version 5	0.8945***	0.7812***	0.5923***	0.5349***	1.0000

Note: Tale-level Pearson correlations among five independent SA–BERT replications of the same prompt for *trust*. Panel A uses raw similarities. Panel B uses residuals from regressions of each run on the Neutral-SA similarity only (the main specification for residualization). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Alignment with an LLM classifier. For external validation, I compare tale-level SA–BERT scores to binary classifications produced by an independent LLM classifier for the same value dimensions. Table II reports unconditional Pearson (point–biserial) correlations, along with correlations after residualizing both measures on Neutral-SA similarity. Correlations remain positive and statistically significant across all dimensions, indicating that convergence between embedding-based and classifier-based measures is not driven by generic folktale register.

Table II: Tale-Level Correlations: SA–BERT vs. GPT

Value	Raw	Adjusted
Honesty	0.154***	0.134***
Trust	0.198***	0.148***
Altruism	0.223***	0.146***
Effort	0.147***	0.091***
Clan	0.170***	0.190***
Ancestor Worship	0.075***	0.112***
Buddhism	0.095***	0.110***
Emperors	0.085***	0.127***
Karma	0.240***	0.221***

Note: Entries are Pearson correlations at the tale level. "Adjusted (Neutral SA)" correlates residuals from regressions of each measure on the Neutral-SA similarity. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Taken together, these exercises indicate that the folktale-based measures capture stable semantic content rather than artifacts of a particular prompt or generic narrative style.

IV Geographic and Historical Correlates of Cultural Themes

A natural question is where the cultural values embedded in folktales come from. Tracing the complete set of origins is beyond the scope of this paper, since norms have been shaped by migration, conquest, market integration, and missionary activity. I therefore focus on a few traits that have played recurrent roles in Chinese social and political history and test a set of hypotheses derived from existing research on ecology, religion, and state formation. The aim is not to provide a comprehensive account of causation but to show that folktale narratives vary systematically with enduring ecological and political conditions.

The first hypothesis links rice agriculture to lineage institutions. Rice cultivation depends on coordinated irrigation and sustained intergenerational labor, whereas wheat cultivation is less cooperative. This ecological contrast is thought to foster collectivist orientations and strengthen kinship organization. Experimental and psychological studies find stronger interdependent mindsets in rice regions of China (Talhelm et al., 2014), and classic ethnography documents powerful lineage institutions in the rice-growing Southeast (Freedman, 1958, 1966). If these social structures encouraged ancestor worship and clan solidarity, counties with high rice suitability should display stronger clan- and ancestor-worship narratives. Columns 1–4 of Table III confirm this expectation, showing a positive association between

rice suitability and the salience of kin-centered themes.

The second hypothesis concerns Buddhism. Confucian orthodoxy, centered on Qufu—the birthplace of Confucius—long stood in tension with Buddhist influence, which spread most successfully where Confucian authority was weaker (Elman, 2013; Gernet, 1995). If the two traditions were substitutes, the prominence of Buddhist stories should increase with distance from the Confucian heartland. Columns 5–6 show precisely this pattern: Buddhist themes are more common in counties farther from Qufu, suggesting that historical variation in religious competition is mirrored in the folktale record.

The third set of hypotheses focuses on the benevolent-emperor ideal, a recurring expression of political legitimacy. One mechanism emphasizes *political proximity*: imperial authority radiated outward from Beijing, the late-imperial capital, implying that narratives celebrating virtuous centralized rule should be stronger near the capital. Columns 7–8 support this prediction. A complementary mechanism emphasizes *hydrological origins of statehood*: the Yellow River basin, especially its Shang–Zhou course, was the cradle of early dynastic power, where collective flood control reinforced hierarchical authority (Pietz, 2021; Chen et al., 2012). If cultural memories of early state formation persisted, benevolent-emperor narratives should be more salient closer to this hydrological core, as confirmed in Columns 9–10.

Taken together, the evidence points to connections between folktale themes and long-run geographic and historical features. Rice suitability is associated with clan and ancestor worship; greater distance from Confucius's birthplace with Buddhist themes; and proximity to Beijing and the Yellow River with the benevolent-emperor motif. While not a full theory of origins, these patterns indicate that regional themes retain ecological and historical imprints that predate modern institutions.

Table III: Geographic and Historical Determinants

Panel A: Clan, Ancestor Worship, and Buddhism

	Clan		Ancestor Worship		Buddhism	
	(1)	(2)	(3)	(4)	(5)	(6)
Rice Suitability	0.0000010*	0.0000013**	0.0000018***	0.0000017***		
	(0.00000058)	(0.00000054)	(0.00000063)	(0.00000058)		
Dist. to Confucius's Birthplace					0.0035**	0.0044***
					(0.0015)	(0.0012)
Latitude	0.000080	-0.000030	0.00012	0.00013	0.00047**	0.00066***
	(0.00021)	(0.00020)	(0.00022)	(0.00020)	(0.00020)	(0.00019)
Longitude	0.00027	0.00040	0.00031	0.00039	0.00016	0.00035
	(0.00030)	(0.00028)	(0.00033)	(0.00030)	(0.00024)	(0.00022)
Population Density 1300	-0.0024***	-0.0015**	-0.0032***	-0.0027***	-0.00065	-0.00087
	(0.00079)	(0.00075)	(0.00085)	(0.00078)	(0.00089)	(0.00084)
Ruggedness	-0.0011*	-0.00080	-0.00086	-0.00072	-0.0011*	-0.00086
	(0.00060)	(0.00057)	(0.00064)	(0.00061)	(0.00062)	(0.00057)
R^2	0.0069	0.0054	0.022	0.025	0.0038	0.011
Obs.	1262	1064	1262	1064	1262	1064

Panel B: Benevolent Emperor

	Benevolent Emperor			
	(7)	(8)	(9)	(10)
Dist. to Beijing	-0.0025**	-0.0024**	-0.0021*	-0.0020*
	(0.0012)	(0.0012)	(0.0012)	(0.0011)
Dist. to Yellow River			-0.0015*	-0.0019***
			(0.00082)	(0.00072)
Latitude	-0.00020	-0.00014	-0.00016	-0.000088
	(0.00017)	(0.00016)	(0.00017)	(0.00016)
Longitude	-0.00039*	-0.00042*	-0.00063**	-0.00071***
	(0.00024)	(0.00023)	(0.00026)	(0.00025)
Population Density 1300	-0.0029***	-0.0024***	-0.0030***	-0.0024***
	(0.00074)	(0.00068)	(0.00073)	(0.00067)
Ruggedness	-0.00020	-0.000080	0.00011	0.00026
	(0.00055)	(0.00051)	(0.00056)	(0.00052)
R^2	0.015	0.015	0.018	0.021
Obs.	1262	1064	1262	1064

Note: Each column reports an OLS regression of a county-level folktale score on the hypothesized geographic determinant with historical/geographic controls. Outcomes are the narrative scores for Clan (Cols. 1–2), Ancestor Worship (3–4), Buddhism (5–6), and the Benevolent Emperor (7–10). The key regressors are: rice suitability (1–4); log distance to Confucius's birthplace, Qufu (5–6); log distance to Beijing (7–8); and log distance to the Yellow River (Shang–Zhou paleochannel), with distance to Beijing included only as a control (9–10). All specifications include population density in 1300, terrain ruggedness, and county coordinates. Odd-numbered columns use the China-proper sample; even-numbered columns restrict to counties with at least two recorded tales. Robust standard errors clustered by county are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$..

V Folklore and Modern Attitudes and Behavior

This section links folktale-based cultural scores—interpreted as the historical imprint of local norms—to individual-level outcomes from the Chinese General Social Survey (CGSS). The outcomes I use are primarily attitudinal measures. For a small number of dimensions where closely aligned attitude questions are not available in the CGSS, I instead rely on the most comparable behavior-based items. By comparing narrative content that plausibly crystallized long before modern development with present-day survey responses, I document systematic associations between past and present cultural traits.

For each core trait—honesty and trust, altruism, effort, and patience—I match CGSS items that capture closely related dimensions, prioritizing attitudes when available.

$$Outcome_{i,c} = \alpha + \beta\ Cultural\ Trait_c + \mathbf{X}_c^H\Omega + \mathbf{X}_c^G\Lambda + \mathbf{X}_c^C\Pi + \mathbf{X}_i\Gamma + \epsilon_{i,c},$$

where i indexes individuals and c counties. $Outcome_{i,c}$ is an individual's CGSS response; $Cultural\ Trait_c$ is the county's folktale score for the matched trait, averaged over five archetype draws. Historical controls ($\mathbf{X}_c^H$) cover agricultural suitability, ruggedness, population density in 1300, treaty-port status, and distance to Confucius's birthplace. Geographic controls ($\mathbf{X}_c^G$) absorb broad macroregion differences. Contemporary controls ($\mathbf{X}_c^C$) include GDP per capita, urbanization, education, sectoral composition, ethnic shares, and an indicator for provincial-capital status. Individual covariates ($\mathbf{X}_i$) include age, gender, and educational attainment. The sample is limited to counties in China proper and excludes counties with large ethnic-minority populations (Figure A.5).

V.A Results

Table IV relates folktale narratives to survey-based measures of honesty, altruism, effort, and patience. Columns 1–4 show that counties with higher folktale scores for honesty and trust report stronger agreement with the statement "People will take advantage of you if you're not careful," indicating alignment between historical narratives praising integrity and reliability and modern expectations of cautious exchange.

Columns 5–6 use a question on whether respondents would intervene if children damaged public property as an index of everyday altruistic concern. Counties rich in altruistic folktales report greater willingness to intervene, consistent with a culture of generosity and communal support.

Columns 7–8 examine agreement with the statement "Personal achievements are largely due to hard work." Counties whose folktales glorify perseverance show stronger endorsement

Table IV: Values and Preferences

	Careful or taken advantage				Stop kids		Success from effort		Save surplus income	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Honesty Narrative	-1.30**	-1.29**								
	(0.54)	(0.54)								
Trust Narrative			-1.60***	-1.63***						
			(0.44)	(0.44)						
Altruism Narrative					1.29***	1.28***				
					(0.39)	(0.39)				
Effort Narrative							8.39**	8.46**		
							(4.10)	(4.18)		
Patience Narrative									1.80**	0.95
									(0.76)	(0.73)
Individual Controls	N	Y	N	Y	N	Y	N	Y	N	Y
Contemporary Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Historical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Mean of Dep. Var.	3.03	3.03	3.03	3.03	2.75	2.75	0.97	0.97	4.09	4.09
R^2	0.011	0.019	0.011	0.019	0.046	0.058	0.074	0.10	0.034	0.066
Observations	21598	21589	21598	21589	4772	4771	4615	4609	4795	4794

Note: This table reports associations between folklore narratives and contemporary attitudes from the Chinese General Social Survey (CGSS). Honesty- and trust-related narratives (cols. 1–4) are linked to responses on whether people will take advantage of you if you are not careful, altruism narratives (cols. 5–6) to whether respondents would stop children from damaging public property, effort narratives (cols. 7–8) to whether success is due to effort, and patience narratives (cols. 9–10) to whether individuals prefer to save surplus income. The explanatory variables measure the average similarity of local folktales to each theme. All regressions include style, contemporary, historical, and geographical controls; individual-level controls are added in even-numbered columns. Columns 7–8 are estimated using logit, and the reported values are pseudo R^2. Standard errors clustered at the county level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of this belief, mirroring narrative emphasis on industriousness and diligence. These models are estimated using logit, and pseudo-R^2 is reported.

Columns 9–10 consider whether individuals would choose to save surplus income. Counties with more patience-themed folktales display stronger tendencies toward thrift and long-term planning in the baseline specification, but the coefficients attenuate and become insignificant once individual-level controls are included. This attenuation is consistent with evidence that patience shapes educational attainment and related individual characteristics captured by these controls, which in turn mediate saving behavior (e.g., Cadena and Keys, 2015; Figlio et al., 2019).

Taken together, variation in folktale content corresponds to systematic differences in survey responses: honesty and trust stories track interpersonal vigilance, altruistic tales track communal engagement, effort tales track endorsement of hard work, and patience tales track

intertemporal preferences. Although moralizing deities are often highlighted as sources of prosocial norms (e.g., Henrich, 2020), these findings suggest that narrative traditions outside formal religious contexts can also capture regional variation in foundational moral values, even when divine authority is not invoked.

Table V: Traditional Beliefs and Practices

	Buddhist affiliation				Ancestor power		Elder care		Kin support		Gift pressure	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Buddhism Narrative	7.00*	6.72*										
	(3.78)	(3.80)										
Karma Narrative			12.5***	12.2***								
			(4.59)	(4.60)								
Ancestor Worship Narrative					-2.73***	-2.33**						
					(1.00)	(0.94)						
Clan Narrative							0.83*	0.74*	2.64***	2.74***		
							(0.48)	(0.41)	(0.92)	(0.93)		
Gift Exchange Narrative											2.64*	2.88**
											(1.34)	(1.34)
Individual Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Contemporary Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Historical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.17	0.19	0.17	0.19	0.063	0.092	0.058	0.079	0.042	0.041	0.085	0.11
Observations	16851	16759	16851	16759	1572	1572	19014	19005	4768	4767	4712	4681
Mean of Dep. Var.	0.053	0.053	0.053	0.053	3.68	3.68	0.53	0.53	3.20	3.20	2.48	2.48

Note: This table reports associations between folktale narrative themes and contemporary beliefs and practices in the CGSS. Columns 1–2 regress self-reported Buddhist affiliation on the Buddhism narrative; columns 3–4 use the Karma narrative as the key explanatory variable for the same outcome; columns 5–6 relate belief in ancestral supernatural powers to the Ancestor-Worship narrative; columns 7–10 link clan-oriented outcomes—responsibility for elder care and willingness to press kin or friends to adopt actions believed beneficial—to the Clan narrative; and columns 11–12 link self-reported financial pressure from gift-giving expenditures to the Gift-Exchange narrative. Even-numbered columns add individual controls (female, age-group, education), and all specifications include contemporary, historical, and geographical controls. Standard errors are clustered at the county level and shown in parentheses. Columns 1–4 are estimated by logit and report pseudo R^2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table V reports associations between folktale narrative themes and contemporary measures of traditional beliefs and practices. Columns 1–2 regress self-reported Buddhist affiliation on the Buddhism narrative, and columns 3–4 use the Karma narrative as the explanatory variable for the same outcome. In both cases, counties whose stories feature stronger Buddhist or karmic content exhibit higher identification with Buddhism.

Columns 5–6 relate belief in ancestral supernatural powers to the prevalence of ancestor-worship narratives. Respondents in counties where such stories are more common are less likely to deny ancestral influence, consistent with stronger adherence to ancestor-related

beliefs.

Columns 7–10 draw on CGSS items concerning responsibility for elder care and willingness to press kin or friends to adopt actions believed beneficial. The estimated coefficients on the clan narrative score are positive across specifications, indicating that stronger clan narratives align with greater endorsement of intergenerational care and kin-based support.

Columns 11–12 link self-reported financial pressure from gift-giving expenditures to the Gift-Exchange narrative. The positive association suggests that in regions where gifting stories are more common, households experience greater financial burden from gift exchange.

Taken together, the results show that historical narratives about Buddhism, karma, ancestor reverence, clan hierarchy, and reciprocal gifting are systematically associated with modern religious affiliation, supernatural belief, family duty, and gifting practice.

Table VI: Trust in State Institutions

	Central Government		Local Government		Military		Police		Judicial System	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Benevolent Emperor Narrative	2.82***	2.06**	1.14	0.50	2.94**	2.39*	3.67**	2.82*	2.05	1.42
	(0.86)	(0.94)	(1.84)	(1.87)	(1.22)	(1.24)	(1.46)	(1.41)	(1.28)	(1.27)
Individual Controls	N	Y	N	Y	N	Y	N	Y	N	Y
Contemporary Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Historical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.095	0.13	0.038	0.061	0.094	0.13	0.077	0.10	0.065	0.085
Observations	4786	4785	4778	4777	4774	4773	4782	4781	4773	4772
Mean of Dep. Var.	4.39	4.39	3.61	3.61	4.37	4.37	3.93	3.93	3.91	3.91

Note: This table reports the association between the prevalence of benevolent-emperor narratives in local folktales and respondents' trust in state institutions. The dependent variables are survey-based measures of trust in the central government (Columns 1–2), local government (3–4), the military (5–6), the police (7–8), and the judicial system (9–10). The main explanatory variable is the county-level average similarity of folktales to benevolent-emperor narratives. All specifications include contemporary, historical, and geographical controls; odd-numbered columns exclude individual controls, while even-numbered columns include them. Standard errors clustered at the county level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VI reports associations between the benevolent-emperor narrative in folktales and contemporary trust in state institutions. Columns 1–2 show a strong positive association with trust in Beijing, consistent with the idea that narratives celebrating virtuous rulers reinforce confidence in the political center.

Columns 3–4 are close to zero, suggesting that imperial stories elevate the legitimacy of the apex of the state without extending to county or township officials.

Columns 5–8 indicate higher trust in the People's Liberation Army and the police in

high-score counties, implying that benevolent-rule narratives also elevate confidence in the coercive arms of the state, particularly when these forces are perceived as guardians of public order on behalf of the central authority.

Columns 9–10 yield null results. Historically, courts were embedded within the bureaucracy rather than functioning as autonomous institutions, and the absence of an effect is consistent with that subordinate status.

Taken together, the benevolent-emperor narrative in folktales is a strong predictor of contemporary trust in the central government, underscoring how historical images of virtuous imperial rule can legitimize today's national institutions. At the same time, the lack of corresponding effects for local governments and courts reflects a well-documented "center–periphery" gap in Chinese public opinion: surveys consistently find high confidence in Beijing but much lower confidence in county and township authorities (e.g., Li, 2004). Where formal local institutions lack credibility, villagers often turn to lineage and temple networks for informal monitoring and public-goods provision (Tsai, 2007), a pattern that aligns with the weak local coefficients observed here.

V.B Robustness

Adding Style Control I account for generic stylistic similarity using the neutral SA control (average similarity to the Nature and Travel neutral archetypes). Including this measure leaves the main results intact. Signs match the baseline in all cases, magnitudes are comparable, and coefficients that are statistically significant in the baseline remain so in most outcomes. The neutral control itself is small and imprecise in most specifications, which suggests that the themed indices are not proxying overall folktale register or cadence. Results are reported in Table A.1.

Without Modern Controls I reestimate the specifications after removing modern covariates while keeping the historical and geographic controls. The associations between the themed indices and the CGSS outcomes remain similar in sign and size, with significance levels that are close to the baseline for the majority of outcomes. This check indicates that the relationships documented in the main tables are not driven by contemporary conditioning variables. Results appear in Table A.2.

VI Cultural Traits and Economic Performance

This section examines the relationship between cultural traits, as derived from folktales, and regional economic outcomes, with a focus on GDP per capita in 2010. All cultural

variables are standardized (z-scored) for comparability. The honesty and trust measures are combined into a single variable, constructed as the average of their standardized values, since they are conceptually too close to separate and empirically highly correlated.

To formally investigate this relationship, I estimate the following equation:

$$\textit{GDP per capita}_c \;=\; \alpha \;+\; \sum_k \beta_k \, \textit{Trait}_{c,k} \;+\; \mathbf{X}_c^C \Gamma \;+\; \mathbf{X}_c^G \Lambda \;+\; \mathbf{X}_c^H \Omega \;+\; \epsilon_c, \tag{1}$$

where $\textit{GDP per capita}_c$ is the log of GDP per capita in county c, and $\textit{Trait}_{c,k}$ represents the set of cultural traits derived from folktales (e.g., honesty, effort, ancestor worship), computed as the average semantic similarity between county c's folktales and the corresponding Synthetic Archetypes, across five iterations of similarity calculations.

$\mathbf{X}_c^C$, $\mathbf{X}_c^G$, and $\mathbf{X}_c^H$ are vectors of contemporary, geographic, and historical control variables, respectively. The set of controls remains consistent with the specification in Section **??**, ensuring comparability, but $\textit{GDP per capita}_c$ is now analyzed as the dependent variable rather than a control. The coefficients β_k capture the marginal effects of cultural traits on GDP per capita, while ϵ_c is the error term.

I implement Lasso regression to identify the subset of cultural traits and control variables most predictive of GDP per capita. This approach penalizes less relevant predictors, enhancing model interpretability and robustness. The analysis includes both a full sample of counties and restricted samples with at least three and at least five folktales, to ensure adequate data coverage for each cultural trait.

Table VII presents the results. Among the most robust findings, *patience* consistently predicts higher GDP per capita, appearing with sizable positive coefficients in both the three-tale and five-tale specifications. This is in line with theories linking patience and intertemporal preferences to investment and growth (Becker, 1964; Doepke and Zilibotti, 2008; Sunde et al., 2022). *Karma* also enters positively across all three samples, including the five-tale specification, though with smaller magnitudes, suggesting that narratives emphasizing karmic justice may reinforce forward-looking or accountability-enhancing behaviors. *Honesty–trust* is positively and significantly associated with GDP per capita in the five-tale specification, the most reliable of the three samples, underscoring the importance of honesty and trust in reducing transaction costs and fostering productivity (Putnam, 1993; Tabellini, 2008; Guiso, Sapienza, and Zingales, 2011). *Effort* appears with positive coefficients in the full and three-tale samples but drops out under the stricter five-tale restriction, making its role less certain but directionally consistent with the importance of industriousness for productivity.

Table VII: Lasso-Selected Cultural Traits Predicting GDP Per Capita (2010)

Cultural Trait	**Any Tale**	**At Least 3 Tales**	**At Least 5 Tales**
Honesty/Trust	–	–	–
Altruism	-0.039	–	–
Effort	0.059	–	–
Patience	–	0.047	0.044
Karma	0.053	0.005	–
Ancestor Worship	-0.075	-0.001	–
Clan	–	-0.008	–
Buddhism	-0.000	0.003	0.007
Benevolent Emperor	–	-0.020	-0.033
Gift Exchange	0.025	–	–

Note: The table reports standardized coefficients from Lasso regressions predicting log GDP per capita in 2010. The first column includes all counties with at least one folktale. The second restricts to counties with at least three folktales, and the third further restricts to counties with at least five folktales. "–" indicates that the corresponding narrative was not selected by the Lasso procedure at the cross-validated penalty parameter. All models include style similarity controls, contemporary conditions, geographic variables, and historical factors.

On the negative side, *clan* narratives are selected in both the three-tale and five-tale samples with negative coefficients, while *ancestor worship* enters negatively in the full and three-tale specifications. These patterns suggest that intensive kinship ties and ritual obligations may limit impersonal exchange and economic dynamism (Enke, 2019; Schulz et al., 2019). Consistent with this interpretation, ancestor veneration in China emphasizes lineage continuity, obligations to forebears, and ritual maintenance—a characteristically backward-looking temporal orientation (e.g., Freedman, 1966; Watson, 1988). *Benevolent emperor* narratives also show negative associations in the three- and five-tale samples, pointing to the economic costs of narratives valorizing centralized authority and hierarchy (Hayek, 1945; Besley and Persson, 2010).

Other traits appear less robust. *Altruism* is negative in the full and five-tale samples, possibly reflecting costly prosocial obligations. *Gift exchange* shows up positively in the full-sample regression but is absent under stricter restrictions, and *Buddhism* does not appear systematically in the selected models.

Taken together, the results highlight patience, karma, and to a lesser extent effort as positive cultural correlates of economic performance, while clan, ancestor worship, and benevolent emperor narratives exhibit negative associations. Other traits such as honesty, altruism, gift exchange, and Buddhism display weaker or sample-specific links. These findings underscore the importance of cultural heterogeneity in shaping long-run economic outcomes, suggesting

that certain values facilitate accumulation and exchange, whereas others embed obligations or political orientations that constrain economic dynamism.

VII Conclusion

This study represents the first systematic effort to construct a cultural map of China by analyzing nearly 20,000 folktales collected in the 1980s. Through this large-scale examination of historical narratives, it uncovers important regional differences in values, mindsets, and beliefs, and links cultural traits—such as trust, honesty, and reciprocity—to contemporary economic outcomes. The results provide robust evidence that culture plays a key role in shaping recent economic growth in China, with traits like honesty, effort, and gift exchange showing strong positive correlations with GDP per capita, while ancestor worship exhibits a consistently negative association.

In addition to contributing to the literature on culture and economics, the findings offer fresh insights into the drivers of China's economic rise by connecting historical cultural narratives to modern economic performance. This perspective underscores the significance of deep-rooted cultural determinants in understanding regional and national growth patterns.

One methodological contribution is that I derive multiple cultural traits from a single, unified data source—Chinese folktales—using a consistent and comparable procedure across traits. This stands in contrast to most existing studies, which typically analyze one cultural dimension at a time and draw on heterogeneous data sources, making cross-trait comparisons less interpretable.

Prior attempts to classify Chinese folklore have been constrained by limited geographic coverage and insufficient regional specificity. By introducing systematic methods for analyzing folktales and their cultural content, this study opens up new opportunities to examine the historical development of cultural traits and their impact on present-day economic and social phenomena. Furthermore, the methodology is readily transferable to other regions with rich cultural archives, providing a valuable framework for exploring the role of cultural narratives in shaping economic outcomes.

References

Algan, Yann, and Pierre Cahuc. 2010. "Inherited Trust and Growth." *American Economic Review* 100 (5): 2060–2092.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100 (401): 464–477.

Barro, Robert J., and Rachel McCleary. 2003. *Religion and Economic Growth.* NBER Working Papers 9682. National Bureau of Economic Research, Inc, May. http://ideas.repec.org/p/nbr/nberwo/9682.html.

Bascom, William. 1965. "Four Functions of Folklore." *The Journal of American Folklore* 67 (266): 333–349.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* Chicago: University of Chicago Press.

Becker, Sascha O, Jared Rubin, and Ludger Woessmann. 2024. "Religion and Growth." *Journal of Economic Literature* 62 (3): 1094–1142.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–1678.

Berezkin, Yu E. 2015. "Folklore and Mythology Catalogue: Its Lay-Out and Potential for Research." *The Retrospective Methods Network,* no. S10, 58–70.

Besley, Timothy. 2020. "State Capacity, Reciprocity, and the Social Contract." *Econometrica* 88 (4): 1307–1335.

Besley, Timothy, and Sacha Dray. 2024. "Trust and State Effectiveness: The Political Economy of Compliance." *Economic Journal* 134 (662): 2225–2251.

Besley, Timothy, and Torsten Persson. 2010. "State Capacity, Conflict, and Development." *Econometrica* 78, no. 1 (January): 1–34.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "The Determinants of Earnings: A Behavioral Approach." *Journal of Economic Literature.*

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

———. 2005. *The Origin and Evolution of Cultures.* Oxford University Press.

Brandt, Loren, and Thomas G Rawski. 2008. *China's Great Economic Transformation.* Cambridge university press.

Cadena, Brian C., and Benjamin J. Keys. 2015. "Human Capital and the Lifetime Costs of Impatience." *American Economic Journal: Economic Policy* 7, no. 3 (August): 126–153.

Chen, Yongqin David, James P. M. Syvitski, Shu Gao, and Albert J. Kettner. 2012. "Socio-economic Impacts on Flooding: A 4000-Year History of the Yellow River, China." *Ambio* 41 (7): 682–698.

Cohn, Alain, Michel A. Maréchal, David Tannenbaum, and Claudia Zünd. 2019. "Civic Honesty Around the Globe." *Science.*

Doepke, Matthias, and Fabrizio Zilibotti. 2008. "Occupational Choice and the Spirit of Capitalism." *The Quarterly Journal of Economics* 123, no. 2 (May): 747–793.

Dundes, Alan. 1980. *Interpreting Folklore.* Bloomington: Indiana University Press.

Eberhard, Wolfram. 1952. *Chinese Festivals.* Great Religious Festivals Series. New York: H. Schuman.

Elman, Benjamin A. 2013. *Civil Examinations and Meritocracy in Late Imperial China.* Harvard University Press.

Enke, Ben. 2019. "Kinship, Cooperation, and the Evolution of Moral Systems." *The Quarterly Journal of Economics* 134 (2): 953–1019.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4): 1645–1692.

Fernández, Raquel. 2011. "Does Culture Matter?" *Handbook of Social Economics* 1:481–510.

Figlio, David, Paola Giuliano, Umut Özek, and Paola Sapienza. 2019. "Long-term orientation and educational performance." *American Economic Journal: Economic Policy* 11 (4): 272–309.

Finnegan, Ruth. 1992. *Oral Traditions and the Verbal Arts: A Guide to Research Practices.* London: Routledge.

Freedman, Maurice. 1958. *Lineage Organization in Southeastern China.* London: University of London, Athlone Press.

———. 1966. *Chinese Lineage and Society: Fukien and Kwangtung.* London: Athlone Press.

Galor, Oded, and Ömer Özak. 2016. "The Agricultural Origins of Time Preference." *American Economic Review* 106 (10): 3064–3103.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57 (3): 535–574.

Gernet, Jacques. 1995. *Buddhism in Chinese Society: An Economic History from the Fifth to the Tenth Centuries.* Translated by Franciscus Verellen. New York: Columbia University Press.

Giuliano, Paola, and Nathan Nunn. 2021. "Understanding Cultural Persistence and Change." *The Review of Economic Studies* 88 (4): 1541–1581.

Golsteyn, Bart H. H., Hans Grönqvist, and Lena Lindahl. 2014. "Adolescent Time Preferences Predict Lifetime Outcomes." *The Economic Journal* 124 (580): F739–F761.

Gombrich, Richard. 2006. *Theravāda Buddhism: A Social History from Ancient Benares to Modern Colombo.* London: Routledge.

Gorodnichenko, Yuriy, and Gerard Roland. 2017. "Culture, Institutions, and the Wealth of Nations." *Review of Economics and Statistics* 99 (3): 402–416.

Greif, Avner, Joel Mokyr, and Guido Tabellini. 2025. *Two Paths to the Twentieth Century: Culture and Institutions in Europe and China, 1000-2000.* Princeton University Press.

Greif, Avner, and Guido Tabellini. 2017. "The Clan and the Corporation: Sustaining Cooperation in China and Europe." *Journal of Comparative Economics* 45 (1): 1–35.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2011. "Civic Capital as the Missing Link." In *Handbook of Social Economics,* edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 417–480. Amsterdam: Elsevier.

Hayek, F.A. 1945. "The Use of Knowledge in Society." *The American Economic Review* 35:519–530.

Heckman, James, and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *American Economic Review.*

Henrich, Joseph. 2020. *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous.* Farrar, Straus / Giroux.

Iannaccone, Laurence R. 1998. "Introduction to the Economics of Religion." *Journal of Economic Literature* 36, no. 3 (September): 1465–1495.

Inglehart, Ronald F. 2018. *Cultural Evolution: People's Motivations Are Changing, and Reshaping the World.* Cambridge: Cambridge University Press.

IPUMS International. 2025. *China 1982: Sample Design and Microdata Documentation.* Minnesota Population Center, University of Minnesota. China 1982 is a 1% systematic household sample. Documentation and sample design materials accessed 2025. https://international.ipums.org/international/sample_designs/sample_designs_cn.shtml.

Jia, Ruixue, and James Kai-sing Kung. 2025. *The Culture and Institutions of Confucianism.* Working Paper, Working Paper Series 33883. National Bureau of Economic Research, May. http://www.nber.org/papers/w33883.

Kanbur, Ravi, and Xiaobo Zhang. 2005. "Fifty Years of Regional Inequality in China: A Journey Through Central Planning, Reform, and Openness." *Review of Development Economics* 9 (1): 87–106.

Keown, Damien. 1992. *The Nature of Buddhist Ethics.* London: Macmillan.

Levi, Margaret. 1988. *Of Rule and Revenue.* University of California Press.

Lévi-Strauss, Claude. 1962. *The Savage Mind.* Originally published in French as *La Pensée Sauvage.* Chicago: University of Chicago Press.

Li, Lianjiang. 2004. "Political Trust in Rural China." *Modern China* 30 (2): 228–258.

Lin, Justin Yifu. 1995. "The Needham Puzzle: Why the Industrial Revolution Did Not Originate in China." *Economic Development and Cultural Change,* 269–292.

Lipset, Seymour Martin. 1959. "Some social requisites of democracy: Economic development and political legitimacy." *The American political science review* 53 (1): 69–105.

Lord, Albert B. 1960. *The Singer of Tales.* Cambridge, MA: Harvard University Press.

Malinowski, Bronislaw. 1926. *Myth in Primitive Psychology.* New York: W. W. Norton Company.

Mauss, Marcel. 1954. *The Gift: Forms and Functions of Exchange in Archaic Societies.* London: Cohen West.

Michalopoulos, Stelios, and Melanie Meng Xue. 2021. "Folklore." *The Quarterly Journal of Economics* (January).

Mokyr, Joel, and Guido Tabellini. 2024. "Social Organizations and Political Institutions: Why China and Europe Diverged." *Economica* 91 (362): 347–382.

Nunn, Nathan. 2012. "Culture and the Historical Process." *Economic History of Developing Regions* 27:S108–S126.

Nunn, Nathan, and Raul Sanchez de la Sierra. 2017. "Why Being Wrong Can Be Right: Magical Warfare Technologies and the Persistence of False Beliefs." *American Economic Review* 107 (5): 582–587.

Nye, John V, and Melanie Meng Xue. 2014. *Raising Dragons.* Working Paper.

Ong, Walter J. 1982. *Orality and Literacy: The Technologizing of the Word.* New York: Methuen.

Perkins, Dwight, et al. 1975. *China's Modern Economy in Historical Perspective.* Stanford University Press.

Pietz, David A. 2021. "The Yellow River, the Chinese State, and the Ecology of North China." In *Oxford Research Encyclopedia of Asian History.* Online edition. Oxford University Press.

Putnam, Robert D. 1993. *Making Democracy Work. Civic Traditions in Modern Italy.* Princeton, NJ: Princeton University Press.

Reimers, Nils, and Iryna Gurevych. 2019. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 3982–3992.

Richardson, Philip. 1999. *Economic Change in China, c. 1800-1950.* Cambridge University Press.

Schulz, Jonathan F., Duman Bahrami-Rad, Jonathan Beauchamp, and Joseph Henrich. 2019. "The Church, Intensive Kinship, and Global Psychological Variation." *Science* 366 (6466): eaau5141.

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

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing like China." *American Economic Review* 101 (1): 196–233.

Sunde, Uwe, Thomas Dohmen, Benjamin Enke, Armin Falk, David Huffman, and Gerrit Meyerheim. 2022. "Patience and Comparative Development." *The Review of Economic Studies* 89 (5): 2806–2840.

Tabellini, Guido. 2008. "Presidental Address Institutions and Culture." *Journal of the European Economic Association* 6 (2-3): 255–294.

Talhelm, Thomas, Xuemin Zhang, Shigehiro Oishi, Chen Shimin, Dongyuan Duan, Xuezhao Lan, and Shinobu Kitayama. 2014. "Large-Scale Psychological Differences Within China Explained By Rice Versus Wheat Agriculture." *Science* 344 (6184): 603–608.

Tsai, Lily L. 2007. *Accountability Without Democracy: Solidary Groups and Public Goods Provision in Rural China.* Cambridge: Cambridge University Press.

Uther, Hans-Jörg. 2004. *The Types of International Folktales: A Classification and Bibliography. Based on the System of Antti Aarne and Stith Thompson.* Vol. 284-286. FF Communications. Helsinki: Academia Scientiarum Fennica.

Vansina, Jan. 1985. *Oral Tradition as History.* Madison, WI: University of Wisconsin Press.

Von Glahn, Richard. 1991. "The Enchantment of Wealth: the God Wutong in the Social History of Jiangnan." *Harvard Journal of Asiatic Studies* 51 (2): 651–714.

———. 2004. *The Sinister Way: The Divine and the Demonic in Chinese Religious Culture.* University of California Press.

Wang, Shaoda, and David Y Yang. 2025. "Policy Experimentation in China: The Political Economy of Policy Learning." *Journal of Political Economy* 133 (7): 000–000.

Watson, Rubie S. 1988. "Remembering the Dead: Graves and Politics in Southeastern China." In *Death Ritual in Late Imperial and Modern China,* edited by James L. Watson and Rubie S. Watson, 121–143. Berkeley: University of California Press.

Wedeman, Andrew H. 2012. *Double Paradox: Rapid Growth and Rising Corruption in China.* Cornell University Press.

White, Cindel J. M., and Aiyana K. Willard. 2025. "How Karma Harms and Helps Generosity Toward Those in Need." *Personality and Social Psychology Bulletin.*

Xue, Melanie Meng. 2024. *High-Value Work and the Rise of Women: The Cotton Revolution and Gender Equality in China.* Working paper.

Yang, Mayfair Mei-Hui. 1994. *Gifts, Favors, and Banquets: The Art of Social Relationships in China.* Ithaca, NY: Cornell University Press.

Zak, P. J., and S. Knack. 2001. "Trust and Growth." *The Economic Journal* 111 (April): 295–321.

Zhu, Xiaodong. 2012. "Understanding China's Growth: Past, Present, and Future." *Journal of Economic Perspectives* 26 (4): 103–124.

Appendix

1 Data

A Geographic Distribution of Folktales

Figure A.1 displays the geographic distribution of all folktales collected by the Folklore Collection Project. Each blue dot represents a county with recorded folktales, and the size of the dot corresponds to the number of tales documented in that county. The map provides a comprehensive visualization of the spatial coverage of folktales across China.

Note that this map includes *all* folktales in the dataset, not just those in the final analytical sample. While the sample used in this study focuses on China proper and regions historically inhabited by the Han Chinese, the figure illustrates the broader geographic scope of the folktale collection effort.

Additionally, some ethnic minority folktales could not be reliably assigned to specific locations, and thus are not shown. Folktales from ethnic minorities that lack clear county-level identifiers are excluded from both the figure and the empirical analysis to ensure consistency in geographic attribution.

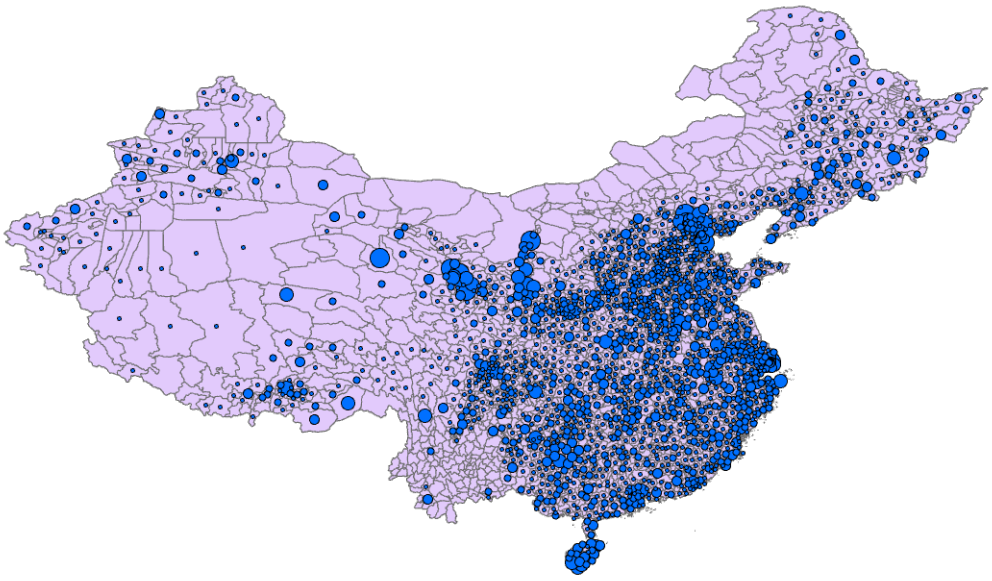
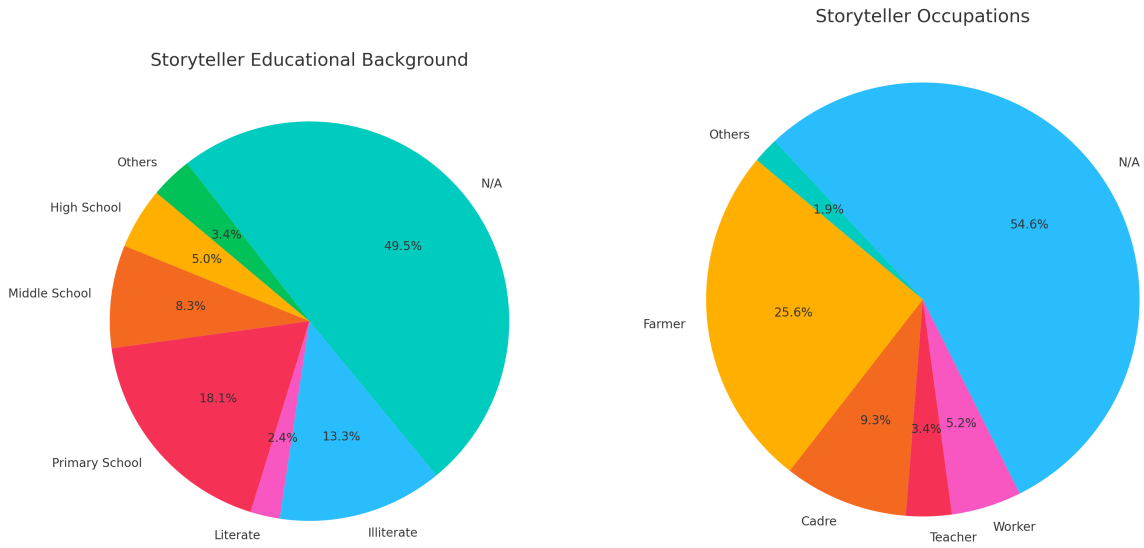


Figure A.1: Geographic Distribution of Folktales

B Background Characteristics of Storytellers



(a) Educational Attainment (b) Occupation Distribution

Figure A.2: Background Characteristics of Storytellers

Note: Panel (a) illustrates the distribution of storytellers by educational attainment, categorized from primary education to college level. Panel (b) presents the occupational composition of storytellers, showing major groups such as farmers, cadres, teachers, and an aggregated "Others" category for minor occupations.

C Ethnic Tales

Figure A.4 displays the number of folktales attributed to each of the top fifteen ethnic minority groups in the original corpus. The Mongol and Hui groups contribute the largest number of tales, followed by Zhuang, Miao, and Tibetan. Groups with smaller representation are aggregated under the "Others" category.

These tales capture a wide range of cultural expressions, many of which diverge significantly from Han narratives in terms of content, values, and narrative structure. For instance, tales from Tibetan, Uyghur, or Kazakh traditions often reflect nomadic life, religious cosmologies, and kinship systems that differ from those prevalent in Han-majority regions. As such, these stories represent valuable cultural materials in their own right and are an important part of China's diverse heritage.

However, the present study restricts its empirical focus to Han-majority regions. Tales associated with ethnic minority groups are not included in the analysis for several reasons. First, many minority folktale collections lack clear geographic identifiers at the county level, making them unsuitable for spatial matching with contemporary economic or attitudinal outcomes. Second, the CGSS survey data used to measure present-day values include only limited coverage of non-Han respondents, which limits the ability to study minority cultural transmission in a systematic way.

The process of sample construction—including the exclusion of ethnic and frontier regions, autonomous areas, and counties where non-Han populations exceed 20%—is described in detail in Section D.

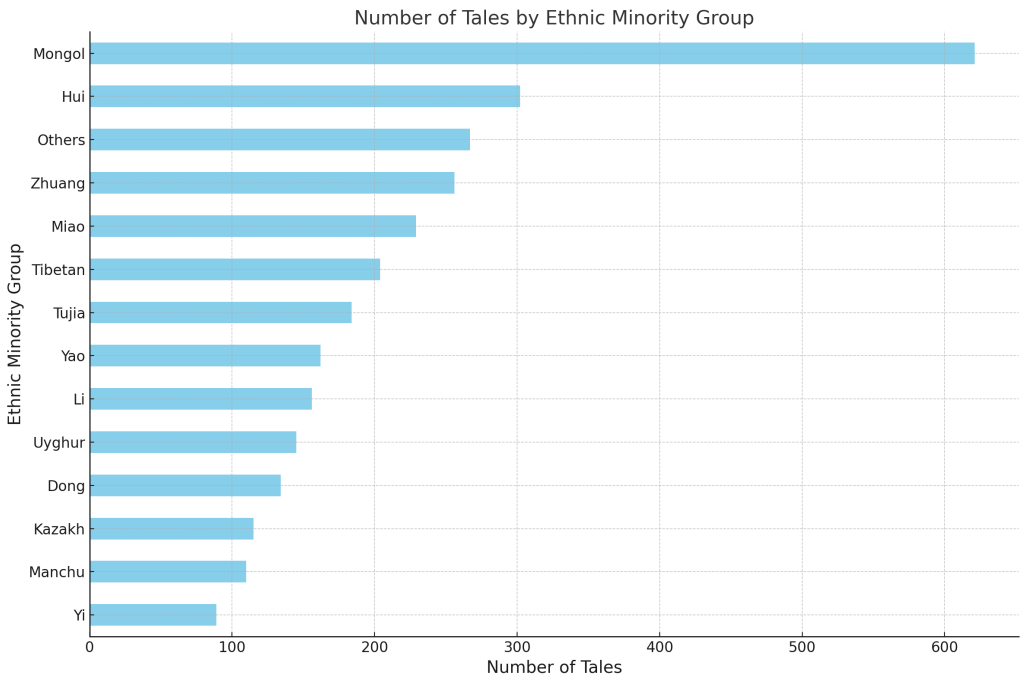


Figure A.4: Number of Folktales by Ethnic Minority Group

Note: This figure displays the number of folktales recorded for each ethnic minority group in the dataset. The Mongol and Hui groups contribute the largest number of tales, followed by groups such as Zhuang, Miao, and Tibetan. The category "Others" includes ethnic groups not listed individually due to smaller tale counts. Only the top fifteen groups by tale count are shown.

D Sample Construction

The original dataset includes folktales from all of China's officially recognized ethnicities. However, this study focuses on regions traditionally occupied by the Han Chinese population. Although the survey data used to measure modern cultural values include some non-Han areas, the vast majority of respondents are Han Chinese, making it difficult to systematically analyze cultural values among ethnic minorities. Furthermore, many folktale collections from minority regions lack clear geographic identifiers, preventing reliable linkage to contemporary outcomes.

The sample is constructed using the following criteria:

1. **Exclusion of Ethnic and Frontier Regions.**
   We exclude provinces and autonomous regions historically or currently recognized as ethnic or frontier areas, such that the sample approximates the Eighteen Provinces of the Qing Dynasty ("China proper"). Territories beyond the Ming Dynasty boundaries (1368–1644) are also excluded.

2. **Exclusion of Ethnic Autonomous Areas.**
   We exclude autonomous prefectures and counties where ethnic populations predominate, as well as regions with legal, cultural, or social norms that have historically diverged from Han Chinese practices.

3. **Population Composition Criterion.**
We exclude counties where more than 20% of the population identified as non-Han in the 2000 census.



Figure A.5: Geographic Distribution of Folktales in the Sample

2 Methodology

A Overview

1. Digitizing Folktales

Original volumes of Chinese folktales were scanned and processed with Optical Character Recognition (OCR) to produce editable text. OCR outputs were manually checked for major recognition errors, and automated scripts were used to normalize punctuation, remove non-story material (e.g., headers, page numbers), and unify common terms to reduce accidental variation in wording.

2. Preparing Synthetic Archetypes

For each cultural theme, five "reference" stories are first identified from the corpus. Candidate stories (about 100 per theme) are screened with a lightweight classifier and then ranked using pairwise model-based comparisons of thematic fit; the top five become exemplars. A separate set of 60 tales—chosen to be stylistically representative of the Han Chinese corpus in length, language, and narrative form—is distilled into a style profile describing common lexical, syntactic, and structural features. This style profile, combined with the theme prompt and the five exemplars, is supplied to a generative model to produce a Synthetic Archetype (SA) for the theme. Each SA includes both the story and a brief explanation of how it matches the theme. The generation is repeated five times to capture minor variation.

3. Computing Semantic Similarity

Semantic similarity is calculated using the `DMetaSoul/sbert-chinese-general-v2` Sentence-BERT model. Each folktale is embedded into a high-dimensional vector space and compared to the five SAs for a given theme; similarities are averaged to yield a folktale–theme score. This process is repeated for all themes.

4. Aggregating by County

Metadata in the collection links each folktale to a county of origin. Folktale–theme scores are grouped by county, and the mean score is taken to represent that county's alignment with the theme. Counties with more stories have proportionally more observations contributing to the mean.

5. Validating Measures

The resulting county-level indices are checked for internal consistency across SA replications, and against control measures that capture only stylistic resemblance. Persistent relationships between the thematic indices and external variables after these controls suggest that the measures are detecting substantive content rather than superficial textual similarity.

B Selection Criteria for the 60 Tales Used in the Style Profile

To create synthetic story styles that accurately reflect the linguistic and structural conventions of Han Chinese folktales, this study relies on a sample of 60 tales drawn exclusively from Han Chinese sources. These tales exclude minority narratives, helping to maintain consistency in language use and cultural context. In order to ensure their suitability for stylistic benchmarking, the average length of the selected tales closely matches that of the broader pool of Han Chinese folktales, so that the sample remains representative in terms of narrative scope and detail.

Random selection further minimizes bias in the choice of tale themes or regional focus, allowing each piece to serve as an authentic illustration of mainstream Han Chinese storytelling. By capturing the underlying patterns of language, plot development, and narration across these 60 tales, the style summary derived from this sample is better positioned to guide the generation of synthetic archetypes. As a result, while no individual story drives the stylistic profile, each contributes distinct features that help to encapsulate the characteristic voice of the folktale corpus.

C Selection Criteria for the Five Reference Tales

For each cultural theme, a set of five reference tales is selected from the Han Chinese folktale corpus to serve as substantive anchors for Synthetic Archetype generation. The selection follows a fixed, multi-stage procedure designed to maximize thematic fit while keeping the reference style aligned with Han folk traditions.

Step 1: Exclusion of Minority Narratives

All tales originating from ethnic minority traditions are excluded to ensure linguistic, stylistic, and cultural consistency with Han folktale norms.

Step 2: Constructing the Candidate Pool

An initial screening is conducted using a lightweight LLM classifier. For a given theme, a tale is included in the candidate pool if any associated sub-prompt receives a rating of "Yes" or "Probably yes." When a theme has multiple sub-prompts, they share the same candidate pool. Each theme's pool contains 100 real tales.

Step 3: Pairwise Thematic Comparisons

All unordered pairs of the 100 candidate tales (4,950 pairs in total) are compared using a model-based evaluation of which tale better fits the specific theme prompt. To avoid forcing a decision when the fit is ambiguous, an option of "Can't decide" is provided.

Step 4: Win Count Ranking

For each tale, the number of "wins" (times it was judged a better thematic fit) is tallied. "Can't decide" responses are neutral and do not contribute to either tale's win count. This produces a ranking of all 100 candidates by their thematic fit score.

Step 5: Selecting the Top Five

The five tales with the highest win counts are chosen as reference exemplars for the theme. These reference tales, together with the 60-tale style profile, are used to condition the generative model in producing Synthetic Archetypes. Each SA thus reflects both the narrative style of the broader corpus and the thematic specificity of the top-ranked exemplars.

D Example Synthetic Archetype: Trust

This appendix reports one of the five Synthetic Archetypes (SAs) generated for the trust theme. County-level trust is constructed by averaging folktale–SA similarity across all five versions, as described in Section II.

> In a remote village lived a kind and gentle elderly woman whom everyone called Granny Li. She was skilled at her craft, and the tofu she made was the best in the village, loved by all. Whenever there was a market day, Granny Li's tofu always sold out quickly, and her stall was full of customers.
>
> One day, a young stranger named Xiao Zhang arrived in the village. Having heard about Granny Li's tofu, he decided to buy some. When he reached her stall, he saw that she was busy and waited patiently. Granny Li looked up and smiled. "Young man, you've been waiting a while, haven't you? This tofu has just been made. I promise it tastes good."
>
> Xiao Zhang nodded. He felt that the old woman was very warm and welcoming, but he hesitated because he only had a few copper coins in his pocket. Granny Li seemed to notice his concern and said with a smile, "It's all right. Take it and eat it first. When you have the money, you can come back and pay me."
>
> Xiao Zhang was surprised and touched. He accepted the tofu and returned home, where he eagerly took a bite. It was indeed delicious. He decided that he would come back the next day to repay her.

The following day, Xiao Zhang returned as promised, bringing a few copper coins. Smiling, he said, "Granny Li, your tofu was wonderful. I've come to pay you back." Granny Li waved her hand and replied, "Young man, the tofu is something I made. Trust is something I give you. The money is not urgent."

Warmth filled Xiao Zhang's heart. He decided to come by often to help, and Granny Li was happy to teach him how to make tofu. Over time, the two became good friends. Xiao Zhang not only learned the craft but later opened a small shop in the village, and his business gradually grew.

Several years later, Xiao Zhang's shop had expanded, and Granny Li's tofu had become a symbol of the village. Xiao Zhang often brought fresh tofu to visit Granny Li, thanking her for the trust she had shown him. Villagers said that Granny Li's trust not only helped Xiao Zhang build a livelihood but also made the village more harmonious.

The story spread throughout the village, and people came to understand the power of trust. Through her honesty and trust, Granny Li fostered cooperation and harmony among others and profoundly changed Xiao Zhang's life.

E Computing BERT Similarity

I use a document–level Sentence-BERT encoder to map each folktale and each synthetic archetype into a fixed-length vector. Specifically, I load `DMetaSoul/sbert-chinese-general-v2` and encode full texts (no sentence segmentation), using GPU with half precision when available for efficiency. Embeddings are L2-normalized and cosine similarity is computed in batches.

Cosine similarity. Let $e(d_i) \in \mathbb{R}^m$ denote the embedding of folktale i and $e(s_{vk})$ the embedding of archetype k for value v. The similarity for the pair (i, v, k) is

$$\mathrm{sim}(d_i, s_{vk}) = \frac{e(d_i) \cdot e(s_{vk})}{\|e(d_i)\| \, \|e(s_{vk})\|} \in [-1, 1].$$

In practice, I precompute $e(s_{vk})$ once per archetype and encode folktales in batches (typically 256 per batch); similarities are clipped to $[-1, 1]$ and written to disk by theme and batch.

County aggregation. For county c, let $\mathcal{D}(c)$ be its folktales. The county–archetype score for value v and replicate k is

$$\mathrm{Trait}_c^{(v,k)} = \frac{1}{|\mathcal{D}(c)|} \sum_{i \in \mathcal{D}(c)} \mathrm{sim}(d_i, s_{vk}).$$

For each value v, I generate $K = 5$ independent archetypes and average the resulting county scores:

$$\mathrm{Trait}_c^{(v)} = \frac{1}{K} \sum_{k=1}^{K} \mathrm{Trait}_c^{(v,k)}.$$

F Neutral Synthetic Archetypes (Style Control)

To account for generic stylistic similarity that is unrelated to the specific cultural themes, I construct three *Neutral Synthetic Archetypes* (SAs): *Nature*, *Travel*, and *Everyday Work*. Prompts reuse the same scaffold and style profile as the themed SAs and request ordinary settings and descriptive action without explicit moral evaluation, religious content, or references to political or social authority. The Everyday Work archetypes depict routine productive activities (such as farming, handicraft, or household work) in a descriptive manner, without emphasizing cooperation, obligation, or moral judgment. For each neutral topic, five independent versions are generated using the same model and decoding settings as in the main pipeline.

Similarity between each tale and each neutral SA version is computed with the Sentence-BERT encoder (mean pooling; L2 normalization) used elsewhere in the paper. For a given topic and version, tale–SA similarities are first averaged to the county. County scores are then averaged across the five versions of that topic, and finally the mean across the three neutral topics is taken to obtain the county-level neutral measure:

$$\text{NeutralSim}_c = \tfrac{1}{3}\Big(\text{Score}_c^{\text{Nature}} + \text{Score}_c^{\text{Travel}} + \text{Score}_c^{\text{Everyday Work}}\Big).$$

The neutral measure enters the regressions as a control at the same level as the themed indices (and the tale-level analogue is used when relevant). Its role is to absorb variation due to shared register, pacing, and narrative cadence so that themed indices are not proxying overall folktale style. Implementation mirrors the main SA pipeline: each Neutral SA consists of a short story in the Han folktale register with a brief rationale; five versions are produced per topic; and similarities are computed and batched with the same encoder and settings as in Sections II – III.

G Random Text Generation and County-Level Text Similarity Control

The random-text control addresses surface-level stylistic similarity arising from local token statistics rather than genre or narrative structure. It complements the Neutral SA control by targeting short-run phrasing and micro-syntax.

Generation. A Markov chain text generator (adapted from `markovify`) is trained on the full folktale corpus with a five-token state. Chinese segmentation uses `jieba`. The resulting "null text" reproduces vocabulary and short-run transitions typical of the corpus but does not form coherent plots or invoke recurring motifs.

Similarity and aggregation. Using the same Sentence-BERT encoder as above, I compute cosine similarity between each folktale and the null text and average these similarities to the county level to form a style-only control, denoted $\textit{StyleControl}_c$. In robustness checks, averaging across multiple independently seeded null texts yields similar results.

Usage. $\textit{StyleControl}_c$ is included alongside the Neutral SA control in specifications that relate themed indices to historical covariates and modern outcomes. The random-text control primarily captures local phrasing and character co-occurrence, while the Neutral SA control captures broader genre features without values. Using both helps separate thematic content from two distinct sources of non-thematic similarity. Diagnostics reported in the appendix

show that correlations with themed indices are modest and that results are stable whether controls are entered individually or together.

3 Cultural Traits: Measures and Validation

A Prompts and Maps

Honesty prompt—The story should mention a character's honest or deceitful behavior, whether such actions drive the plot or influence relationships. Specifically, the narrative should either show how a character wins trust, respect, or cooperation through steadfast honesty—highlighting its positive role in personal growth, interpersonal relationships, and social order.

Altruism prompt—The story should include acts of charity—such as donating, assisting, or giving alms—regardless of whether the outcome is straightforward or complex. Specifically, the narrative should demonstrate how charitable actions improve the recipient's life and promote social harmony and warmth.

Effort prompt—The story should mention a character's efforts or struggles, irrespective of whether these lead to success or failure. Specifically, the narrative should depict how perseverance and hard work enable the character to overcome obstacles and achieve goals, emphasizing effort as essential for personal development and social achievement.

Patience prompt—The story should show a character who exercises patience and strategic foresight—enduring setbacks, waiting for the right moment, or sustaining effort over many years—before ultimately succeeding. The plot should highlight how long-run planning, delayed gratification, or steady investment in learning, work, or relationships yields rewards that short-term gains cannot match.

Buddhism prompt

Karma prompt—The story should revolve around the Buddhist law of cause and effect—"good is rewarded, evil is punished." It should depict characters who, when choosing between wrongdoing and virtuous conduct (or when suffering injustice), personally witness the cycle that "as one sows, so shall one reap." The plot must portray karmic loops in which generosity, repentance, rescuing others, or refusing to harm eventually resolve the crisis and transform the protagonist's heart—often catalyzed by the appearance of Kṣitigarbha, a vision of Guanyin, or a monk reciting the *Sutra of Cause and Effect*. Readers should see a clear, positive transformation brought about by karmic retribution.

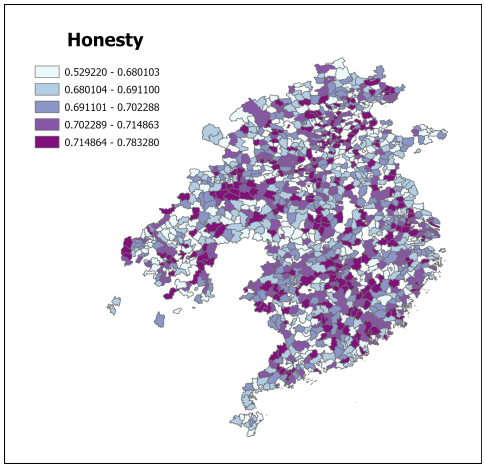
Ancestor worship prompt—The story should portray characters' reverence and remembrance of their forebears, showing how ancestors—seen as symbols of wisdom, virtue, and strength—shape descendants' beliefs, morals, and life choices. It should highlight how ancestor worship sustains family continuity, prompting characters to uphold honor, follow ancestral teachings, pass on skills, or fulfill ancestral missions. In adversity, ancestral protection or legacy should provide spiritual support, strengthening intergenerational obligations.

Clan organization prompt—The story should describe the rules and hierarchical structures within a lineage, emphasizing members' obedience to clan authority. It should show how characters follow elders' directives, protect family honor, or fulfill lineage responsibilities—even when this entails personal sacrifice or moral conflict.

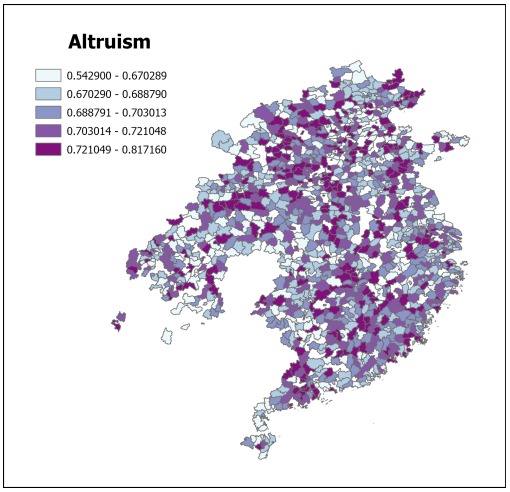
Gift exchange prompt—The story should center on giving and reciprocation of gifts,

showing how this practice fosters bonds, expresses respect, or fulfills obligations within the community. It should demonstrate how reciprocal gifting reinforces social ties, maintains harmony, or signals social status.

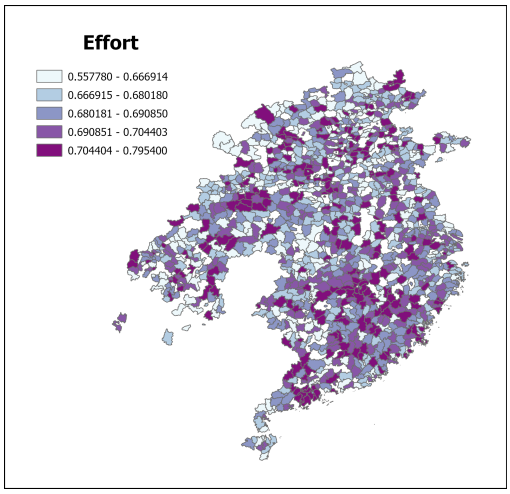
Benevolent emperor prompt —The story should depict how a wise, enlightened emperor governs with benevolence, justice, and sound judgment, ensuring social stability, economic prosperity, and the well-being of the people. It should showcase how the emperor's decisions gain public support and maintain lasting order, presenting him as an ideal sovereign whose governance enhances the welfare of the entire empire.



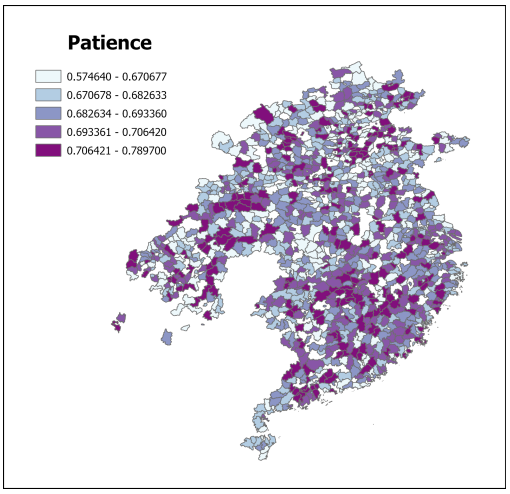
(a) Honesty



(b) Altruism

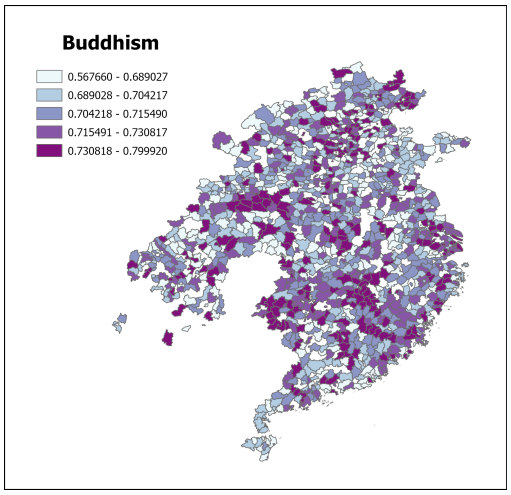


(c) Effort

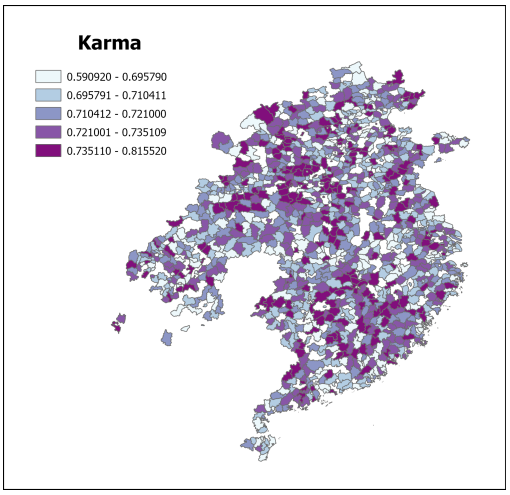


(d) Patience

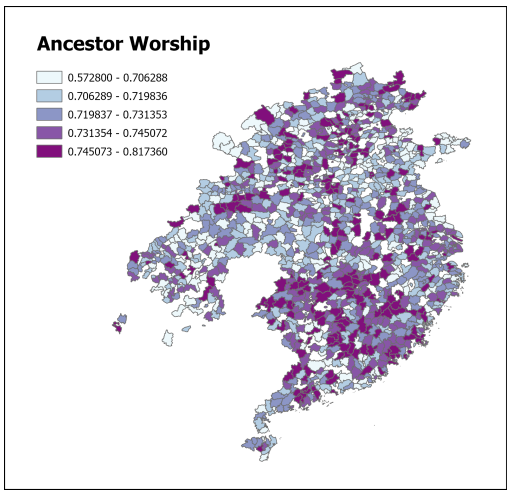
Figure A.6: County-Level Cultural Traits from Folktales



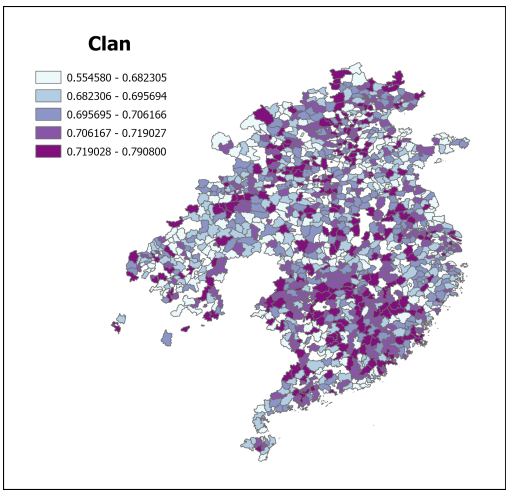
(e) Buddhism



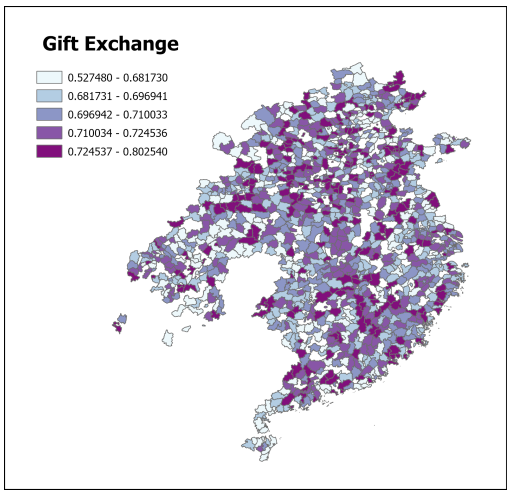
(f) Karma



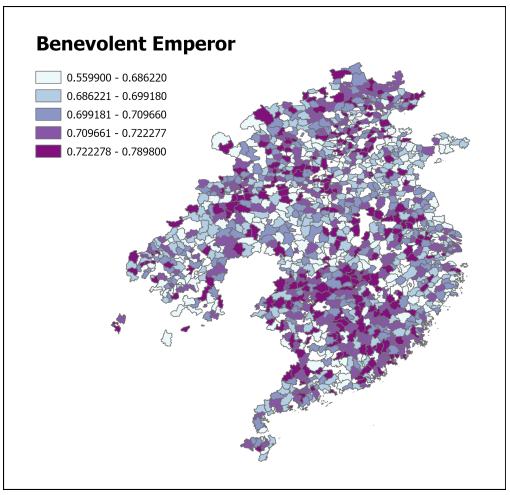
(g) Ancestor Worship



(h) Clan



(i) Gift Exchange



(j) Benevolent Emperor

Figure A.6: County-Level Cultural Traits from Folktales (continued)

Note: Each panel maps the county-level SA–BERT similarity measure for the indicated trait. Within each panel, darker shading indicates stronger semantic alignment between local folktales and the corresponding Synthetic Archetypes.

Table A.1: Robustness: Folktale Narratives and Contemporary Outcomes (Neutral SA Control)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Honesty Narrative	-1.47**															
	(0.57)															
Trust Narrative		-1.95**														
		(0.91)														
Altruism Narrative			2.11***													
			(0.76)													
Effort Narrative				8.02												
				(5.04)												
Patience Narrative					1.47											
					(1.22)											
Buddhism Narrative						6.96*										
						(4.02)										
Karma Narrative							7.68*									
							(4.67)									
Ancestor Worship Narrative								-4.70***								
								(1.43)								
Clan Narrative									0.70	2.65**						
									(0.53)	(1.01)						
Gift Exchange Narrative											3.15					
											(2.59)					
Benevolent Emperor Narrative												3.16**	0.77	2.63*	3.39*	1.93
												(1.53)	(2.29)	(1.42)	(1.75)	(1.59)
Neutral SA (Travel & Nature)	0.032	0.79	-0.51	2.06	0.60	4.52	2.90	2.12**	0.099	-0.74	-1.42	0.57	1.37	0.36	0.72	0.55
	(0.45)	(0.84)	(0.87)	(5.39)	(0.96)	(4.32)	(4.29)	(1.04)	(0.28)	(1.00)	(2.15)	(0.88)	(1.03)	(0.85)	(0.84)	(0.87)
Contemporary Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Historical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2 / Pseudo R^2	0.012	0.013	0.026	0.063	0.030	0.15	0.15	0.059	0.074	0.038	0.082	0.093	0.033	0.080	0.071	0.073
Observations	24410	24410	5448	5251	5470	18991	18991	1806	21554	5445	5386	4351	5452	5440	5454	5444
Mean of Dep. Var.	3.05	3.09	2.74	0.97	4.07	0.053	0.053	3.67	0.51	3.18	2.45	4.34	3.62	4.36	3.90	3.87

Notes: Each column reports the association between a county-level folktale-based SA–BERT measure (tale–SA similarity aggregated to the county) and a CGSS outcome, controlling for contemporary, historical, and geographic covariates. All specifications include a style-neutral control (average similarity to non-thematic SAs: travelogue and nature). Outcomes by column: (1–2) prudence vs. being taken advantage of (honesty) and interpersonal trust; (3) willingness to stop children from taking advantage of others (altruism); (4) belief that success comes from effort—logit (effort); (5) propensity to save surplus income (patience); (6–7) Buddhist affiliation—logit (Buddhism, karma); (8) belief in ancestral supernatural power (ancestor worship); (9–10) clan-related outcomes: elder care performed by children; pressure from kin/friends (clan); (11) perceived gift-giving pressure (gift exchange); (12–16) trust in state institutions—central government, local government, military, police, courts (benevolent emperor). Standard errors clustered by county. Pseudo R^2 is reported for logit columns. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Robustness: Folktale Narratives and Contemporary Outcomes (Excluding Modern Controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Honesty Narrative	-2.00***															
	(0.61)															
Trust Narrative		-1.82**														
		(0.77)														
Altruism Narrative			1.06*													
			(0.55)													
Effort Narrative				8.46+												
				(5.60)												
Patience Narrative					1.16											
					(1.10)											
Buddhism Narrative						13.4***										
						(4.78)										
Karma Narrative							5.16									
							(4.86)									
Ancestor Worship Narrative								-1.36*								
								(0.75)								
Clan Narrative									1.12**	3.07**						
									(0.46)	(1.42)						
Give Exchange Narrative											2.48+					
											(1.53)					
Benevolent Emperor Narrative												2.63+	1.19	2.86**	4.18**	2.52+
												(1.62)	(1.56)	(1.34)	(1.62)	(1.63)
Historical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geographical Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2 / Pseudo R^2	0.010	0.012	0.020	0.042	0.015	0.12	0.12	0.047	0.055	0.020	0.050	0.078	0.027	0.054	0.047	0.049
Observations	24410	24410	5448	5251	5470	18991	18991	1806	21554	5445	5386	4351	5452	5440	5454	5444
Mean of Dep. Var.	3.05	3.09	2.74	0.97	4.07	0.053	0.053	3.67	0.51	3.18	2.45	4.34	3.62	4.36	3.90	3.87

Notes: Each column reports the association between a county-level folktale-based SA–BERT measure (tale–SA similarity aggregated to the county) and a CGSS outcome. Specifications exclude modern controls and retain historical and geographic covariates. All specifications include the style-neutral control based on non-thematic SAs (travelogue and nature). Outcomes by column follow the main specification: (1–2) prudence vs. being taken advantage and interpersonal trust; (3) willingness to stop children from taking advantage of others; (4) belief that success comes from effort — logit; (5) propensity to save surplus income; (6–7) Buddhist affiliation — logit; (8) belief in ancestral supernatural power; (9–10) clan-related outcomes; (11) perceived gift-giving pressure; (12–16) trust in state institutions — central government, local government, military, police, courts. Standard errors clustered by county. Pseudo R^2 is reported for logit columns. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.