# Values of China: Toward a Cultural Map*

Melanie Meng Xue†

LSE

March 2026

**Abstract**

This paper constructs a county-level cultural map of China using nearly 20,000 folktales collected nationwide in the 1980s. I introduce a synthetic-archetype approach that combines narrative prompts with sentence-level BERT embeddings to measure how strongly local folktale corpora reflect value- and social-order narratives, and I aggregate tale-level scores within counties to obtain county-level measures by theme. Reliability is assessed using stability across independent archetype draws, and text-similarity controls help separate thematic content from narrative form. Using these measures, I document systematic historical and geographic correlates and relate the folktale-based measures to contemporary attitudes in the Chinese General Social Survey, including trust and prosocial behavior, beliefs about effort and intertemporal orientation, religion, kin obligations, gift pressures, and views of political authority. In a prediction-oriented screen for county log GDP per capita in 2010 using cross-validated LASSO with rich historical, geographic, and contemporary controls, economic performance is most strongly associated with narratives emphasizing effort, patience, and gift exchange on the positive side and ancestor worship and benevolent central authority on the negative side.

**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, Jennifer Pan, Louis Putterman, Moses Shayo, Uwe Sunde, Guido Tabellini, Clemence Tricaud, Nico Voigtländer, 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, Toronto, UCLA Anderson, and Utrecht University, as well as participants 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), Hoover Conference on Text-as-Data in Policy Analysis (2026), KCL Workshop in Comparative Politics and Political Economy (2024), and 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 led, for their help during the early experimentation stage of the project, when the work relied primarily 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 important for economic growth and political stability, both in historical settings such as the Industrial Revolution and in modern development (Tabellini, 2008a; Gorodnichenko and Roland, 2017; Besley, 2020). This perspective aligns with research emphasizing how societal beliefs and narratives shape long-run economic choices, social norms, and institutional structures (Bénabou and Tirole, 2006; Fernández, 2011). A related survey-experimental literature shows that policy preferences depend not only on interests, but also on beliefs about how policies work and what they imply for efficiency and redistribution (Stantcheva, 2021; Kuziemko et al., 2015).

Comparative analyses of cultural and institutional evolution in Europe and China have also provided influential accounts of divergent developmental pathways (Greif and Tabellini, 2017; Mokyr and Tabellini, 2024). For instance, Greif and Tabellini (2017) emphasizes how China's clan-based social organization shaped collective decision-making and hierarchy, relative to European settings where individualistic frameworks were more compatible with the emergence of democratic governance and market-based economies. Similarly, Mokyr and Tabellini (2024) highlights how differences in social organization affected innovation and long-run economic progress.

Yet we have far less systematic evidence on how specific belief- and value-related traits vary within China at fine geographic levels, and on how to measure those traits consistently across places. More generally, there is no widely used framework for surveying many cultural traits at once within countries in a way that is comparable across traits and locations. Much of the existing evidence comes either from a narrow set of survey items or from studies of particular practices in isolation, which makes it difficult to compare dimensions, study how they co-vary, or relate them to a common set of outcomes across places.

China is also a setting where standard sources are especially incomplete for this purpose. Historical barriers to anthropological and field-based documentation during much of the socialist period, together with the limited geographic granularity of widely used modern surveys such as the CGSS, make it difficult to recover a county-level picture of locally embedded beliefs and practices from conventional data.

These gaps are particularly striking given the historical continuity of cultural traits in

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

China. China combines deep historical records and diverse local traditions with relatively recent and uneven urbanization, so many locally embedded practices plausibly persisted well into the late twentieth century. Scholars such as Perkins et al. (1975) and Richardson (1999) have highlighted the importance of enduring local practices in shaping governance, economic behavior, and social interactions. Historical narratives, captured in folktales and other cultural artifacts, provide a useful lens for studying 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.[1] Together, these similarity measures provide a quantitative summary of how strongly each county's narrative corpus reflects each cultural theme.

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 geographically disaggregated measures of beliefs and attitudes in China using historical narratives. By aggregating thematic content from nearly 20,000 folktales, this research constructs a county-level map of embedded value- and belief-related narratives, capturing dimensions of local moral reasoning and social expectations that are typically difficult to observe at scale. This map fills a significant gap in the literature, where within-country analyses of heterogeneity in beliefs and attitudes often rely on a narrow set of observable measures across locations and domains.[2] The approach here covers a wider set of belief domains and leverages vernacular folktales to recover locally embedded attitudes that can be linked to contemporary economic and political

[1] For background on text-as-data methods in economics, see Gentzkow, Kelly, and Taddy (2019). For an application to China that uses LLM-based coding of policy documents to construct interpretable indices, see Fang, Li, and Lu (2025).

[2] Pan and Xu (2017) provide a rare exception in the political domain, using a large-scale online survey to study the configuration of ideological preferences in China and its geographic and socioeconomic correlates.

outcomes.

Second, the methodological framework used in this research contributes to a growing literature on cultural quantification. While much empirical work proxies culture using surveys and closely related measures of preferences, beliefs, and norms, this study shows how qualitative narrative sources can be systematically analyzed to construct comparable measures of cultural themes across space.[3] The paper uses synthetic reference archetypes together with BERT-based semantic similarity to extract thematic content from a large folktale corpus. By combining these tools with transparent construction and aggregation rules, the approach yields geographically disaggregated measures that are straightforward to replicate and extend. This methodology can be applied to other narrative corpora and settings, enabling comparative studies of cultural persistence and its economic implications.

Third, the research introduces historical cultural traits as an additional explanatory factor for economic outcomes within 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; Xu, 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 effort and patience, 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 growth experience and its uneven regional development.

Fourth, this study contributes to the economics of religion and supernatural beliefs by bringing new evidence from China on non-monotheistic and locally specific belief systems, as reflected in vernacular narratives. While much existing work focuses on monotheistic traditions, this study highlights ancestor worship, Buddhist influences, and supernatural motifs embedded in folklore. 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.[4] Prior research shows that religion can shape incentives and social

[3]Michalopoulos and Xue (2021) develop an approach that uses ConceptNet to map folktale motifs into interpretable concept categories and construct measures of cultural content.

[4]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.

organization (Iannaccone, 1998; Barro and McCleary, 2003; Becker, Rubin, and Woessmann, 2024), and that supernatural beliefs can support cooperation through informal social enforcement (Nunn and Sierra, 2017). Related historical scholarship documents how supernatural figures such as Wutong encoded moral ambivalence about wealth and commerce (Von Glahn, 1991, 2004), and other work shows that culturally specific beliefs can affect behavior in measurable ways (Nye and Xue, 2014). By quantifying these belief-related themes in folktales, the paper provides evidence on their spatial variation and their association with contemporary attitudes and regional economic development.

Finally, the study contributes to the long-run development literature by using within-country evidence from China to assess the contemporary relevance of historically embedded beliefs and values. 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. By providing systematic evidence on cultural heterogeneity within a single country and linking it to contemporary outcomes, this study complements cross-country analyses and adds evidence on the long-run relevance of historically embedded beliefs and values.

The remainder of this paper is organized as follows. Section II describes the data and the methodological framework used to construct county-level cultural variables from folktales. Section III defines the main traits and reports validation and robustness exercises for the folktale-based measures. Section IV documents geographic and historical correlates of selected cultural themes. Section V relates the folktale-based measures to contemporary outcomes in the CGSS. Section VI explores the relationship between cultural traits and county-level economic performance. Section VII 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 structure 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 rea-

soning and social values that are difficult to recover from modern surveys (Malinowski, 1926; Lévi-Strauss, 1962; Michalopoulos and Xue, 2021).

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. Folklore-based measurement can recover interpretable cultural content from narrative sources, making it a useful complement to survey-based approaches (Michalopoulos and Xue, 2021).

### *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. Because the collections were built through interviews, storytellers were typically older rural residents, a group for whom exposure to nationally standardized content was often limited. In these cohorts,

literacy was low and, in many dialect regions, command of national-standard Mandarin (Putonghua) was often limited, which lowered effective exposure to national print and broadcast programming and reduced the scope for homogenization of local narrative material relative to later decades.[5] 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.

In addition, the narrative setting of the tales is informative about when the underlying material likely took shape. Folktales come from long-standing oral traditions, so precise dating is difficult. However, many narratives are set in historical China rather than the post-1949 period. Using contextual cues in the story setting and institutions, I classify 41.1% of tales as set in pre-1949 China, including 21.5% that are clearly pre-twentieth century. Only 0.48% are set in the post-1949 era, while 36.9% are of indeterminate period. This distribution suggests that a substantial share of the recorded narratives draw on cultural material formed prior to the institutional and ideological transformations of the PRC period.

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 refrain from rewriting narratives into standardized forms. The guidance emphasized faithful transcription even when the material was morally uncomfortable or repugnant by contemporary standards, on the view that sanitizing content would distort what was being documented. Folklore research in this period therefore prioritized documentation over editorial improvement, with the aim of preserving the narratives in the form in which they were recounted rather than producing harmonized or nationally polished 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 la-

[5] 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 60 are coded as illiterate. See the IPUMS International documentation for the China 1982 sample design and variable definitions (IPUMS International, 2025).

bor 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 spans all provinces in China and documents 18,228 tales. Each province contributed a dedicated volume, yielding a geographically organized archive. The number of tales per province ranges from 335 to 1,128.

Most tales can be linked to specific counties. Out of the 18,228 total tales, 17,070 are assigned to a county, so nearly 94% of the archive can be geocoded at the county level. The remaining observations lack a unique county identifier because the published record attributes them to an ethnic minority group or to a broader location label rather than a specific county. These observations are therefore not used when constructing county-level measures.

The archive covers a wide range of narrative types, including myths, legends, local stories, and morality tales. Examples include *Pangu Separates Heaven and Earth* and *The Butterfly Lovers*. Tale length varies substantially. The mean length is about 981 words.

**Storytellers and Collectors** The compilation of folktales drew on local storytellers, many of whom worked in agriculture in rural China. The average age of storytellers is 60.4 years. In the tale-level metadata, 2,885 tales are attributed to storytellers coded as illiterate and 4,352 tales are attributed to storytellers with "elementary or below" schooling, indicating only rudimentary literacy and very limited formal schooling. Tales attributed to storytellers in agriculture account for 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 across storytellers.

Collectors who recorded the tales tend to have more formal schooling and are younger on average, with a mean age of 40.7 years. In the tale-level metadata, 277 tales are recorded by collectors with "elementary or below" schooling, 6,451 tales are recorded by collectors with

secondary schooling (junior high, senior high, or vocational), and 4,088 tales are recorded by collectors with college or above. Only 28 tales are recorded by collectors coded as illiterate. 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** The corpus includes tales attributed to ethnic minority groups. I set these aside in the baseline analysis and report descriptive statistics for them in Appendix A.3.

### *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.[6] 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 style of Chinese folktales. Using Synthetic Archetypes rather than observed tales as reference texts reduces measurement noise and improves comparability across space. A single historical tale inevitably bundles the target concept with idiosyncratic plot elements, characters, and local phrasing, so similarity to that tale can reflect story-specific features rather than the underlying value. Synthetic Archetypes are generated to isolate the core narrative logic of each theme while remaining stylistically consistent with the corpus, which makes cosine similarity more likely to capture semantic alignment with the concept of interest rather than resemblance to a particular story. Using multiple archetypes per theme further stabilizes the measure by reducing sensitivity to any single reference narrative, following standard reliability logic for composite measures formed by averaging multiple parallel items (Cronbach, 1951).

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,

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

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 within counties to form county-level measures 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, 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 provides an overview of the cultural measures constructed from folktales. I organize them into three domains based on the conceptual content they target and on how prompts are framed.

## *III.A Overview of the 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 measures differ in the conceptual domain they target and in how prompts are framed.

**Values and Preferences** The first domain 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; Sunde et al., 2022).

For this domain, 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 C.

**Traditional Beliefs and Practices** The second domain 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 an ethic that links present intent to future consequences, fostering compassion, honesty, and self-discipline where secular enforcement is weak. 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 C.

**Legitimacy of Central Authority** The third domain captures beliefs about the legitimacy of centralized state authority. A long tradition in political economy and political sociology emphasizes that when state authority is viewed as rightful and trustworthy, compliance is easier to sustain, enforcement costs fall, and the state can more effectively provide public goods (Lipset, 1959; Levi, 1988; Besley and Dray, 2024). A recurring theme in the folktale corpus is the benevolent emperor, a ruler portrayed as just and welfare-improving. I focus on this theme because it provides a concrete way to measure beliefs about rightful central authority using the same text-based pipeline as the other cultural families.

I therefore construct a county-level benevolent-emperor measure. Higher values indicate that local folktales more often draw on this image of virtuous central rule. Appendix C reports the prompt, the corresponding Synthetic Archetype, and the county-level map.

### *III.B Worked Example: Patience*

To make the construction concrete, I illustrate the procedure using *patience*. I begin with a fixed prompt template and generate five independent Synthetic Archetype (SA) stories. I then embed each folktale and each SA with the same Chinese Sentence-BERT model and compute cosine similarity between every folktale and each SA. For each folktale, I average its similarity across the five SAs, and I then aggregate these tale-level scores to the county level.

> **Patience prompt.** Stories should describe how characters succeed through patience, waiting for the right moment, or sustaining long-term effort. The plot should emphasize the positive role of patience in learning, careers, social relationships, or life planning.

Two features of this prompt are important for interpretation. First, it encodes a positive framing requirement, patience must be portrayed as beneficial and rewarded rather than merely mentioned. Second, averaging across independently generated SAs reduces dependence on any single archetype realization and improves the stability of the resulting county measure.

Figure I presents the resulting county-level patience measure.

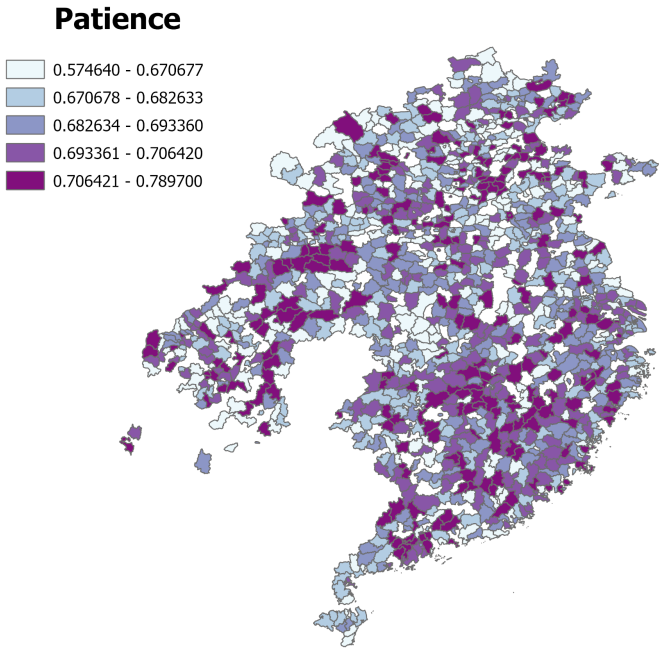



Figure I: County-Level Patience Measure

Note: The figure shows the county-level SA–BERT measure for patience. For each folktale, I compute cosine similarity to each of five independently generated patience SAs and average across the five; I then aggregate the resulting tale-level scores to the county level. The legend bins counties into five quantiles of the county distribution. Darker shading indicates higher values, meaning that local folktales more often portray patience as beneficial and rewarded.

### *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. Appendix C.3 separately checks that the Synthetic Archetype narratives themselves correspond to the intended themes. 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. Neutral-SA similarity serves as a style control, absorbing generic resemblance across tales that is unrelated to the target value (Appendix B.5). Residualizing on Neutral-SA similarity partials out

variation due to shared pacing, narrative cadence, and other common stylistic features. High correlations in both panels indicate that the extracted signal is stable to stochastic variation in SA generation and is not driven by generic story 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). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

**Coherence with an independent LLM classifier.** As an external coherence check, I compare tale-level SA–BERT scores to classifications produced by an independent LLM classifier for the same value dimensions. The classifier is prompted with the same conceptual definitions used to generate the Synthetic Archetypes, but it applies them directly to each tale and returns a four-category judgment: "yes," "probably yes," "probably no," or "no." I code "yes" and "probably yes" as a positive classification and treat the remaining responses as negative.[7] Table II reports unconditional Pearson (point–biserial) correlations between

[7]In practice, the LLM classifications are noisy and appear over-inclusive for some themes, with the share of positive classifications reaching roughly 40% in the most permissive cases. The exercise is therefore best interpreted as an external validation check: it asks whether an independently produced set of labels covaries with the embedding-based scores, rather than treating the classifier rates as a precise estimate of prevalence. The positive correlations reported in Table II support such alignment.

the SA–BERT scores and the resulting binary classifier outcomes, along with correlations after residualizing both measures on Neutral-SA similarity. Correlations remain positive and statistically significant across all dimensions, indicating that alignment between embedding-based scores and classifier-based labels is not driven by generic narrative style.

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*** |
| Patience | 0.205*** | 0.142*** |
| Buddhism | 0.095*** | 0.110*** |
| Ancestor Worship | 0.075*** | 0.112*** |
| Clan | 0.170*** | 0.190*** |
| Gift Exchange | 0.194*** | 0.187*** |
| Benevolent Emperor | 0.085*** | 0.127*** |

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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

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

## IV Geography, History, and Values in Folktales

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 relies on coordinated irrigation, synchronized labor inputs, and durable village-level cooperation, whereas wheat cultivation is less irrigation-intensive and can be organized with weaker local coordination. This ecological contrast is thought to foster more interdependent social norms and strengthen kin-based 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 reinforced collective ritual and kin-based sanctioning, 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.

Table III: Geographic and Historical Determinants

| | Clan | Ancestor Worship | Buddhism | Benevolent Emperor | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Rice Suitability | 0.109* | 0.173*** | | | |
| | (0.063) | (0.063) | | | |
| Dist. to Confucius's Birthplace | | | 0.094** | | |
| | | | (0.040) | | |
| Dist. to Beijing | | | | -0.091** | -0.078* |
| | | | | (0.044) | (0.044) |
| Dist. to Yellow River | | | | | -0.089* |
| | | | | | (0.049) |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mean of Dep. Var. | 0.70 | 0.73 | 0.71 | 0.70 | 0.70 |
| $R^2$ | 0.0069 | 0.022 | 0.0038 | 0.015 | 0.018 |
| Observations | 1262 | 1262 | 1262 | 1262 | 1262 |

Note: Note: Each column reports an OLS regression of a county-level folktale score on the hypothesized geographic determinant with historical/geographic controls. Standardized beta coefficients are reported. Outcomes are the narrative scores for Clan (column 1), Ancestor Worship (column 2), Buddhism (column 3), and the Benevolent Emperor (columns 4–5). The key regressors are rice suitability (columns 1–2), log distance to Confucius's birthplace, Qufu (column 3), log distance to Beijing (column 4), and log distance to the Yellow River (Shang–Zhou paleochannel), with log distance to Beijing included only as a control (column 5). Controls include log population density in 1300, log ruggedness, latitude, and longitude. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The second hypothesis concerns Buddhism. Buddhism entered China in the late Han through Central Asian networks and caravan routes, and early communities formed around translation activity and court-linked patronage in North China. Over the subsequent centuries, especially during the era of political fragmentation, Buddhism expanded geographically and adapted to local social and intellectual life, eventually becoming a major institutional presence with monasteries, clerical networks, and periodic state support (Zürcher, 1959; Ch'en, 1964; Gernet, 1995). At the same time, Confucian statecraft remained the idiom of official governance, and the later imperial examination system helped consolidate a Confucian moral and textual canon (Elman, 2013).

This diffusion history motivates a simple implication for spatial variation. If Confucian

authority was strongest near its symbolic heartland around Qufu, Buddhist influence may have been comparatively constrained there, while gaining greater salience where Confucian institutional and cultural pull was weaker. Under this substitutability logic, the prominence of Buddhist stories should increase with distance from Qufu. Columns 5–6 show precisely this pattern. Buddhist themes are more common in counties farther from Qufu, consistent with the idea that long-run religious competition and uneven diffusion are mirrored in the folktale record.

The third set of hypotheses focuses on the benevolent-emperor ideal, a recurring expression of political legitimacy in which social order is attributed to virtuous centralized rule. One mechanism emphasizes *political proximity*. If imperial authority and its associated moral language were more salient near the capital—through administrative presence, elite circulation, and the visibility of central institutions—then narratives celebrating virtuous rule should be stronger closer to Beijing, the late-imperial 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 central to early dynastic power, where collective flood control and large-scale hydraulic management 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. Columns 9–10 confirm this pattern, consistent with enduring spatial variation in the cultural representation of political legitimacy.

The results link folktale themes to long-run geographic and historical features in ways consistent with existing hypotheses. 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. These examples illustrate that the folktale measures capture durable cultural differences connected to ecological conditions and political history.

## V Folklore and Modern Attitudes and Behavior

This section links folktale-based cultural measures—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 county-level alignment with archetype tales that reflect the target traits to 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-based semantic similarity 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.

### *V.A Results*

Table IV relates folktale narratives to survey-based measures of honesty, altruism, effort, and patience. Columns 1–4 use the CGSS item "People will take advantage of you if you are not careful" (higher values indicate stronger agreement). Counties with higher honesty-affirming and trust-affirming folktale scores report *lower* agreement with this statement (negative coefficients), consistent with less interpersonal suspicion in places where folktales emphasize integrity and reliability.

Columns 5–6 use a question on whether respondents would stop children from damaging public property as a measure of everyday prosocial engagement. Counties with higher altruism-affirming folktale scores report a greater willingness to intervene.

Columns 7–8 examine whether respondents agree that success is mainly due to effort. Counties whose folktales place greater emphasis on effort are more likely to endorse this belief. These models are estimated using logit, and pseudo-$R^2$ is reported.

Columns 9–10 consider whether respondents report that, "If I have extra money, I first consider saving it." Counties with more patience-valuing folktales display stronger saving-first tendencies in the baseline specification, but the coefficient attenuates and becomes statistically insignificant once individual-level controls are included. Appendix D.1 provides additional evidence by adding age-group and education controls in steps.

Overall, variation in folktale content corresponds to systematic differences in conceptually

Table IV: Values and Preferences

| | Careful or taken advantage | | | | Stop kids | | Success from effort | | Save first | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Honesty-affirming | -1.30** | -1.29** | | | | | | | | |
| | (0.54) | (0.54) | | | | | | | | |
| Trust-affirming | | | -1.60*** | -1.63*** | | | | | | |
| | | | (0.44) | (0.44) | | | | | | |
| Altruism-affirming | | | | | 1.29*** | 1.28*** | | | | |
| | | | | | (0.39) | (0.39) | | | | |
| Effort-valuing | | | | | | | 8.39** | 8.46** | | |
| | | | | | | | (4.10) | (4.18) | | |
| Patience-valuing | | | | | | | | | 1.80** | 0.95 |
| | | | | | | | | | (0.76) | (0.73) |
| Individual Controls | | ✓ | | ✓ | | ✓ | | ✓ | | ✓ |
| Contemporary Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Historical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Geographical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 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 folktale-based cultural measures and contemporary attitudes in the Chinese General Social Survey (CGSS). The dependent variables are responses to the following items. Columns 1–4 "People will take advantage of you if you are not careful." Columns 5–6 whether respondents would stop children from damaging public property. Columns 7–8 whether success is mainly due to effort. Columns 9–10 whether respondents prefer to save surplus income. The explanatory variables are county-level averages of semantic similarity between local folktales and the corresponding Synthetic Archetypes, oriented so higher values indicate more pro-trait narrative content. All specifications include style, contemporary, historical, and geographic controls. Odd-numbered columns exclude individual controls, while even-numbered columns include them (sex, age group, education). Columns 7–8 are estimated by logit and report pseudo $R^2$; other columns are estimated by OLS and report $R^2$. Standard errors clustered at the county level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

related survey responses. Honesty- and trust-affirming narratives are associated with lower agreement that others will take advantage, altruistic narratives are associated with reported intervention in everyday settings, effort narratives are associated with endorsement of hard work as a driver of achievement, and patience narratives are associated with intertemporal orientation. Although moralizing deities are often emphasized as sources of prosocial norms (e.g., Henrich, 2020), these patterns suggest that vernacular narrative traditions can also capture local variation in moral content even when divine authority is not invoked.

Table V links the folktale-based cultural measures to contemporary beliefs and practices in the CGSS. Columns 1–2 use a binary indicator for Buddhist affiliation. Counties with higher Buddhism scores have higher rates of self-reported Buddhism.

Columns 3–4 use the item on belief in ancestors' supernatural power, coded so that lower

Table V: Traditional Beliefs and Practices

| | Buddhist affiliation | | Ancestor power | | Kin support | | Gift pressure | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Buddhism | 7.001* | 6.778* | | | | | | |
| | (3.59) | (3.54) | | | | | | |
| Ancestor Worship | | | -3.027*** | -2.813*** | | | | |
| | | | (0.96) | (0.92) | | | | |
| Clan | | | | | 2.338** | 2.417*** | | |
| | | | | | (0.90) | (0.90) | | |
| Gift Exchange | | | | | | | 2.225+ | 2.492* |
| | | | | | | | (1.38) | (1.38) |
| Individual Controls | | ✓ | | ✓ | | ✓ | | ✓ |
| Contemporary Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Historical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Geographical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mean of Dep. Var. | 0.0534 | 0.0535 | 3.668 | 3.668 | 3.176 | 3.176 | 2.450 | 2.453 |
| Pseudo $R^2$ / $R^2$ | 0.153 | 0.174 | 0.0577 | 0.0815 | 0.0377 | 0.0369 | 0.0822 | 0.105 |
| Observations | 18991 | 18971 | 1806 | 1806 | 5445 | 5443 | 5386 | 5351 |

Note: This table reports associations between folktale-based cultural measures and contemporary beliefs and practices in the Chinese General Social Survey (CGSS). Columns 1–2 use self-reported Buddhist affiliation as the dependent variable. Columns 3–4 use belief in ancestral supernatural powers. Columns 5–6 use willingness to press kin or friends to adopt actions believed beneficial. Columns 7–8 use self-reported financial pressure from gift-giving expenditures. The explanatory variables are county-level averages of semantic similarity between local folktales and the corresponding Synthetic Archetypes, oriented so higher values indicate more pro-trait narrative content. All specifications include style, contemporary, historical, and geographic controls. Odd-numbered columns exclude individual controls, while even-numbered columns include them (sex, age group, education). Columns 1–2 are estimated by logit and report pseudo $R^2$; columns 3–8 are estimated by OLS and report $R^2$. Standard errors clustered at the county level are shown in parentheses. $^{+}$ $p < 0.15$, $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.

values indicate stronger belief (1 = believe, 2 = neither, 3 = do not believe). The coefficient on the ancestor-worship measure is negative, which implies stronger belief in ancestral influence in higher-score counties.

Columns 5–6 use agreement with the statement "I will try to get my relatives and friends to do things that I think are beneficial to them, even if they do not agree." The estimated coefficients on the clan measure are positive, indicating stronger endorsement of kin-based persuasion and intra-network influence in higher-score counties.

Columns 7–8 relate reported financial pressure from gift expenditures to the gift-exchange measure. The positive association indicates higher reported gift-related pressure in counties with stronger gift-exchange narratives.

In sum, the Buddhism, ancestor worship, clan, and gift-exchange measures are systematically related to matched survey outcomes on religious affiliation, belief in ancestral power,

kin-oriented influence, and gifting pressure.

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 | 2.626*** | 1.866* | 1.756 | 1.133 | 2.893** | 2.310* | 3.912** | 3.116** | 2.323* | 1.767 |
| | (0.87) | (0.95) | (2.00) | (2.03) | (1.27) | (1.29) | (1.50) | (1.45) | (1.31) | (1.29) |
| Individual Controls | | ✓ | | ✓ | | ✓ | | ✓ | | ✓ |
| Contemporary Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Historical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Geographical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mean of Dep. Var. | 4.366 | 4.366 | 3.615 | 3.616 | 4.355 | 4.355 | 3.900 | 3.901 | 3.872 | 3.872 |
| $R^2$ | 0.0875 | 0.131 | 0.0331 | 0.0562 | 0.0799 | 0.123 | 0.0709 | 0.0977 | 0.0726 | 0.0927 |
| Observations | 5461 | 5459 | 5452 | 5450 | 5440 | 5438 | 5454 | 5452 | 5444 | 5442 |

Note: This table reports associations between benevolent-emperor narratives in local folktales and respondents' trust in state institutions in the Chinese General Social Survey (CGSS). 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 explanatory variable is the county-level average semantic similarity between local folktales and the Benevolent-Emperor Synthetic Archetype, oriented so higher values indicate more benevolent-emperor narrative content. All specifications include style, contemporary, historical, and geographic controls. Odd-numbered columns exclude individual controls, while even-numbered columns include them (sex, age group, education). 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 measure in local folktales and contemporary trust in state institutions. Columns 1–2 show a strong positive association with trust in the central government, consistent with the idea that cultural traditions emphasizing virtuous rulers are associated with greater confidence in the political center.

Columns 3–4 are close to zero, indicating that this relationship does not extend in the same way to county or township officials.

Columns 5–10 show positive associations with trust in the People's Liberation Army, the police, and the judicial system in high-score counties, suggesting that the benevolent-emperor measure is also associated with confidence in major state institutions beyond the central government.

Overall, the benevolent-emperor measure is a strong predictor of contemporary trust in the central government. The contrast between central and local government trust aligns with a well-documented center–local 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), which is

consistent with the weaker 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, Travel, and Everyday Work neutral archetypes). Including this control leaves the main results intact. Signs match the baseline throughout, coefficient magnitudes are comparable, and coefficients that are statistically significant in the baseline remain so for most outcomes. The neutral control itself is typically small and imprecisely estimated, suggesting that the themed measures are not primarily driven by generic stylistic similarity. Results are reported in Table A.2.

**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.3.

## 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:

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

where $\textit{GDP per capita}_c$ is the log of GDP per capita in county $c$, and $\textit{Cultural Trait}_{c,k}$ 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 V,

ensuring comparability, but *GDP per capita*$_c$ is now analyzed as the dependent variable rather than a control. The coefficients $\beta_k$ capture the marginal effects of cultural traits on GDP per capita, while $\epsilon_c$ is the error term.

I use cross-validated LASSO to screen which folktale-based cultural measures are most predictive of county log GDP per capita, conditional on the full set of historical, geographic, and contemporary covariates used elsewhere in the paper. This exercise is prediction-oriented rather than causal. The appeal of LASSO in this setting is that it selects a sparse subset of predictors from a relatively rich candidate set, which helps limit overfitting and yields a more interpretable list of traits that add incremental predictive content (Tibshirani, 1996; Hastie, Tibshirani, and Wainwright, 2015). Table VII reports the standardized coefficients on the folktale measures selected at the cross-validated penalty parameter.

Table VII: LASSO screen: Folktale-based predictors of log GDP per capita

| | (1) Baseline | (2) Reduced folklore set | (3) + Text similarity | (4) + Neutral SA + text sim. |
|---|---|---|---|---|
| *Selected folktale-based predictors (standardized coefficients)* | | | | |
| Ancestor Worship | −0.0644 | −0.0652 | −0.0682 | −0.0634 |
| Effort-valuing | +0.0372 | +0.0375 | +0.0432 | +0.0379 |
| Gift Exchange | +0.0278 | +0.0279 | +0.0204 | +0.0239 |
| Patience-valuing | +0.0169 | +0.0169 | +0.0095 | +0.0125 |
| Benevolent Emperor | −0.0024 | −0.0024 | −0.0026 | |
| Trust-affirming | | +0.0004 | | |
| Buddhism | | | +0.0069 | +0.0066 |
| *Cross-validation and fit* | | | | |
| Selected $\lambda$ | 0.0016821 | 0.0015326 | 0.0013965 | 0.0013965 |
| Out-of-sample $R^2$ | 0.5443 | 0.5445 | 0.5440 | 0.5438 |
| Observations | 1,211 | 1,211 | 1,211 | 1,211 |

Note. Each column reports standardized coefficients on folktale-based measures among the predictors selected by cross-validated LASSO in a county-level regression of log GDP per capita. Column 1 uses the full baseline set of folktale measures, together with the historical/geographic and contemporary covariates used elsewhere in the analysis. Column 2 uses a reduced folklore set that keeps only one measure from each closely related pair in the taxonomy, retaining the trust-affirming narrative rather than honesty-affirming. Column 3 additionally allows a generic text-similarity proxy to enter as a style control. Column 4 allows both text similarity and Neutral SA similarity to enter the candidate set. Reported $R^2$ values are out-of-sample measures from the cross-validation procedure for the full candidate set in each specification.

Across specifications, the selected folklore measures enter with stable and economically interpretable signs. Effort and patience are selected with positive coefficients, consistent with evidence linking forward-looking preferences and self-control to income and development outcomes across settings (Becker and Mulligan, 1997; Galor and Özak, 2016; Falk et al., 2018). Gift exchange also enters with a positive sign, in line with work emphasizing reciprocity and informal enforcement as supports for exchange when formal contracting is limited (Greif, 1993; Fafchamps, 2004; Tabellini, 2008b). Ancestor worship enters with a negative sign,

consistent with arguments that dense kin obligations can concentrate cooperation within the lineage and narrow the scope for broader impersonal exchange that supports market integration (Alesina and Giuliano, 2015; Enke, 2019). Out-of-sample fit is essentially unchanged across columns, suggesting that these screens are not driven by a particular modeling choice.

Column 2 implements a restricted screen that prunes the folklore candidate set by keeping only one measure from closely related concepts, retaining the trust-affirming narrative rather than honesty-affirming. Under this restriction, trust is selected with a positive sign, consistent with the large literature associating generalized trust with higher income and faster growth (Knack and Keefer, 1997; Guiso, Sapienza, and Zingales, 2004; Algan and Cahuc, 2010). The qualitative conclusions for the other selected folklore predictors remain the same.

The benevolent-emperor measure continues to enter when included, with a small negative sign. This is consistent with the interpretation motivating the measure, namely that benevolent-emperor narratives encode a paternalistic view of political order in which welfare and protection are expected from a vertical relationship rather than from horizontal coordination and institutional constraint. In environments where long-run prosperity depends on impersonal exchange and institutions that limit arbitrary power, such orientations can be associated with lower income (North, 1990; Acemoglu, Johnson, and Robinson, 2001; Acemoglu and Robinson, 2012). Allowing generic text-similarity proxies to enter the candidate set in Columns 3 and 4 leaves the main selected folklore measures and their signs largely intact, which is consistent with the folklore signals reflecting substantive content rather than generic stylistic similarity.

# VII Conclusion

This paper represents the first systematic effort to construct a within-China cultural map that measures multiple cultural traits in a unified framework and relates them to contemporary economic and political outcomes. Using a common semantic pipeline across concepts, I recover systematic regional differences in values, beliefs, and practices from the local folktale corpus. Two findings stand out. First, these measures track modern attitudes and behaviors in survey data in predictable ways. Second, a subset remains informative for income differences across counties even after conditioning on rich historical, geographic, and contemporary covariates.

The survey evidence links historical folktale content to present-day attitudes and behaviors. Counties with higher honesty and trust scores are associated with stronger agreement with caution-in-exchange items. Counties with higher altruism scores display greater will-

ingness to intervene in everyday public-good settings, and counties with higher effort scores exhibit stronger endorsement of hard work as a driver of achievement. For patience, the association with thrift and saving behavior is strongest in the baseline specification and attenuates once individual-level controls are included, consistent with patience operating partly through education and related individual characteristics.

The within-China income results are consistent with the view that durable local traits can matter for development outcomes. In the GDP analysis, measures tied to effort, patience, and gift exchange enter with positive signs in multivariate screens, while ancestor worship and the benevolent-emperor measure enter with negative signs. These patterns summarize which dimensions of the folktale-based measures remain informative in a multivariate setting.

More broadly, the findings also connect to a longstanding historiography on whether China's modern transformation should be understood primarily as a response to external influence or whether it also drew on domestically rooted sources of change.[8] While this paper does not attempt to adjudicate between these views, the results are consistent with the idea that locally embedded beliefs, as reflected in vernacular narratives, are economically relevant in the modern period and are systematically associated with the geography of development within China.

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.

[8] A classic statement of the "impact and response" framing is Teng and Fairbank (1979). For a critique and a call for more China-centered approaches in modern Chinese historiography, see Cohen (1984).

# References


Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review* 91, no. 5 (December): 1369–1401.

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail.* New York: Crown Business.

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

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, and Casey B Mulligan. 1997. "The Endogenous Determination of Time Preference." *The quarterly journal of economics* 112 (3): 729–758.

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.

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.

Ch'en, Kenneth K. S. 1964. *Buddhism in China: A Historical Survey.* Princeton, NJ: Princeton University Press.

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.

Cohen, Paul A. 1984. *Discovering History in China: American Historical Writing on the Recent Chinese Past.* New York: Columbia University Press.

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

Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16 (3): 297–334.

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.

Fafchamps, Marcel. 2004. *Market Institutions in Sub-Saharan Africa: Theory and Evidence.* Cambridge, MA: MIT Press.

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.

Fang, Hanming, Ming Li, and Guangli Lu. 2025. *Decoding China's Industrial Policies.* Working Paper 33814. National Bureau of Economic Research, May. Accessed February 1, 2026. https://www.nber.org/papers/w33814.

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. Originally published in French as *Les aspects économiques du bouddhisme dans la société chinoise du Ve au Xe siècle* (Saigon: École française d'Extrême-Orient, 1956). New York: Columbia University Press.

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

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

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Trader's Coalition." *American Economic Review* 83 (3): 525–548.

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. 2004. "The Role of Social Capital in Financial Development." *American Economic Review* 94, no. 3 (June): 526–556.

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations.* Boca Raton, FL: CRC Press.

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.

Knack, Stephen, and Philip Keefer. 1997. "Does social capital have an economic payoff? A cross-country investigation." *The Quarterly journal of economics* 112 (4): 1251–1288.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments." *American Economic Review* 105 (4): 1478–1508.

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* 1:54.

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

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance.* Cambridge: Cambridge University Press.

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.

Pan, Jennifer, and Yiqing Xu. 2017. "China's Ideological Spectrum." *Journal of Politics.*

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.

Stantcheva, Stefanie. 2021. "Understanding Tax Policy: How Do People Reason?" *The Quarterly Journal of Economics* 136 (4): 2309–2369.

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

———. 2008b. "The Scope of Cooperation: Values and Incentives." *The Quarterly Journal of Economics* 123, no. 3 (August): 905–950.

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.

Teng, Ssu-yü, and John King Fairbank. 1979. *China's Response to the West: A Documentary Survey, 1839-1923.* Harvard University Press.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–288.

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.

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

Xu, Chenggang. 2011. "The Fundamental Institutions of China's Reforms and Development." *The Journal of Economic Literature* 49 (4): 1076–1151.

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.

Zürcher, Erik. 1959. *The Buddhist Conquest of China: The Spread and Adaptation of Buddhism in Early Medieval China.* Vol. 11. Sinica Leidensia. Leiden: E. J. Brill.

# For Online Appendix

# A Data

## *A.1 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.

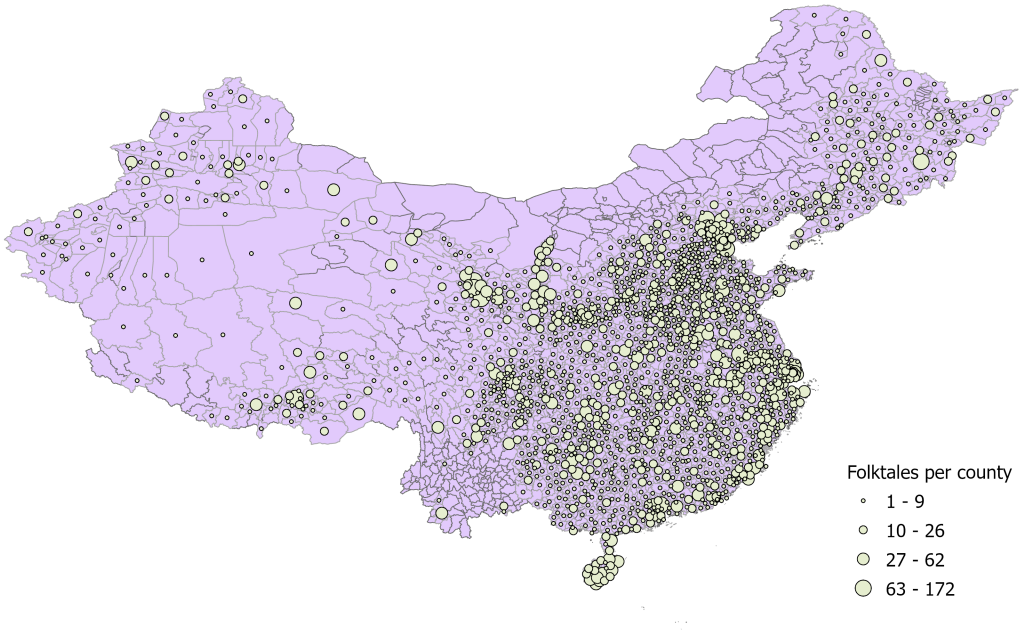
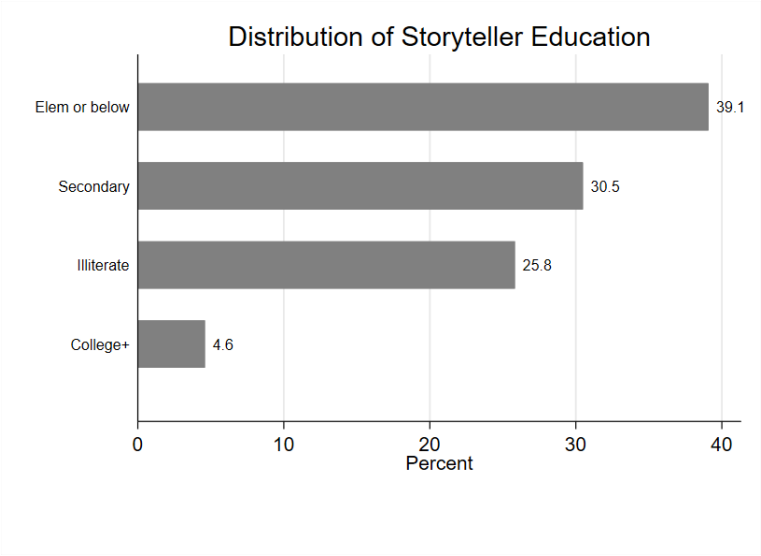


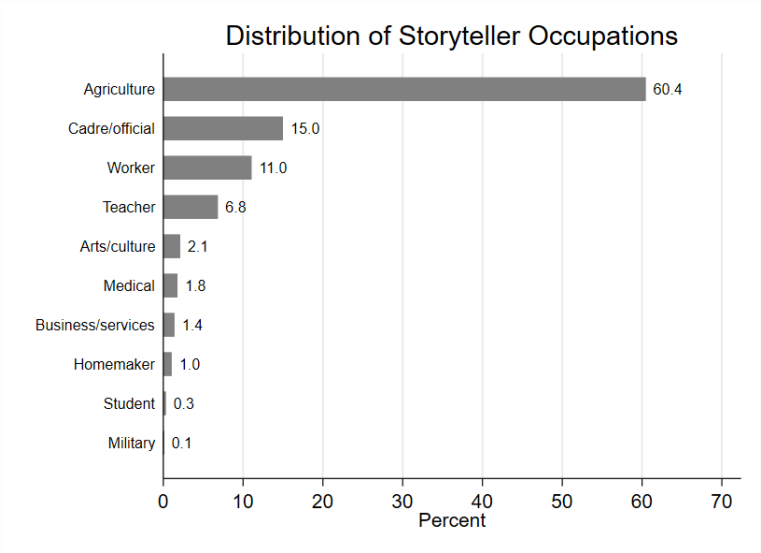

Figure A.1: Geographic Distribution of Folktales

Note: This map shows the total number of folktales in the corpus that can be linked to each county. Circles are placed at county centroids and sized by the number of linked tales; the legend reports the underlying bins. Counties with no circle are counties for which no folktale can be linked in the available metadata. County boundaries are shown for geographic reference.

## *A.2 Background Characteristics of Storytellers*




(a) Storyteller Education




(b) Storyteller Occupation

Figure A.2: Background Characteristics of Storytellers

Note: Panel (a) reports the distribution of storytellers' education categories and panel (b) reports the distribution of storytellers' occupations. The underlying metadata are recorded at the tale level; for this figure, I collapse the data to unique storytellers and compute percentages across storytellers. Some storytellers are excluded because their education or occupation is missing or too unstructured to code; in panel (b), some exclusions reflect occupations that fall outside the ten-category taxonomy.

### *A.3 Ethnic Tales*

Figure A.4 summarizes the number of folktales by recorded ethnic minority group in the metadata and displays the top 20 groups by tale count.

These ethnic minority tales are an important part of the original corpus. In the baseline analysis, I set them aside because the empirical design relies on county-level spatial matching and on validation using contemporary survey outcomes, and both steps are less consistently feasible for the minority collections. In particular, a non-trivial share of minority tales is recorded without stable county identifiers, which makes linkage to county-level covariates and outcomes unreliable. In addition, the survey data used for validation provide limited coverage of minority respondents at fine geographic levels, limiting statistical power and comparability across groups and space.

The sample-construction rules that implement these exclusions are described in Appendix A.4. A systematic analysis would require consistent subprovincial identifiers for minority collections and survey data with sufficient coverage to validate value measures within minority communities. These conditions are not met in the present data.

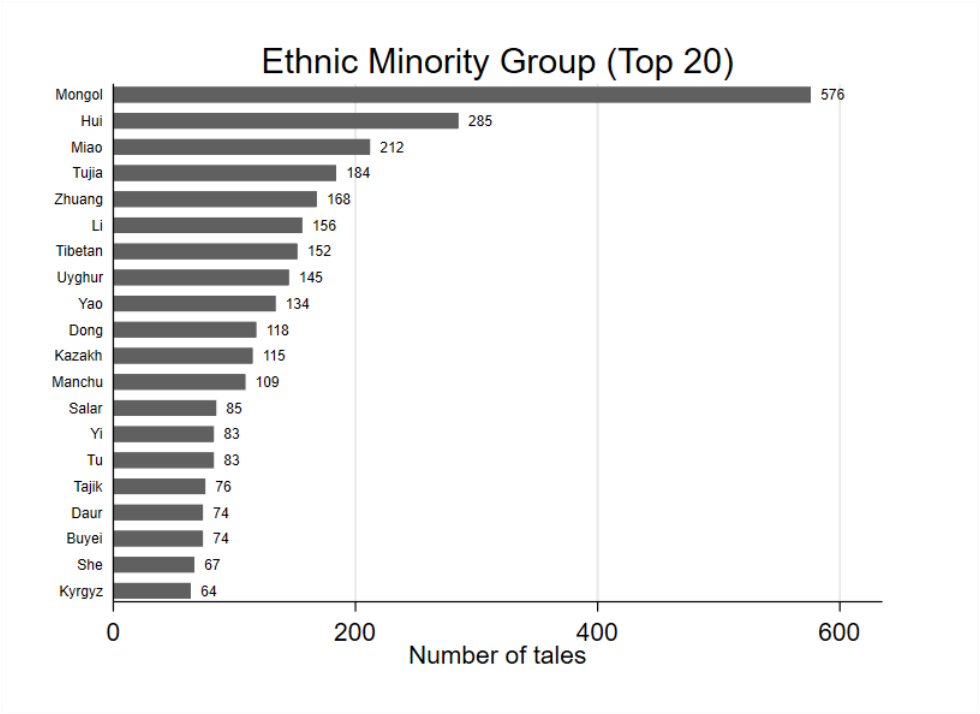



Figure A.4: Ethnic Minority Group of Tales

Note: This figure reports the number of folktales by recorded ethnic minority group in the metadata. Bar heights are tale counts. The figure displays the top 20 groups by count. Tales without a recorded minority label are excluded; in the source metadata, ethnicity is typically recorded only when a specific minority group is indicated.

### *A.4 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.

# B Methodology

## *B.1 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.2 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.

### *B.3 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.

### *B.4 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:

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

### *B.5 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.

Sentence-BERT maps each text into a vector representation, and cosine similarity between a tale and an SA measures semantic proximity in that representation. Neutral SAs provide a summary measure of a tale's generic similarity to common folktale writing conventions. Including Neutral-SA similarity as a control improves interpretability by absorbing variation that is shared broadly across stories and is not specific to the paper's target dimensions.

Similarity between each tale and each neutral SA version is computed with the same Sentence-BERT encoder used elsewhere in the paper (mean pooling; L2 normalization). 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:

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

Implementation mirrors the main SA pipeline. Each Neutral SA is a short story written to resemble the corpus in length and narrative form and includes 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.

### *B.6 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 $StyleControl_c$. In robustness checks, averaging across multiple independently seeded null texts yields similar results.

**Usage.** $StyleControl_c$ 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.

# C Cultural Traits: Measures and Validation

## *C.1 Prompts*

**Trust prompt**—The story should describe how characters build or maintain interpersonal trust (or cope with suspicion) in ways that promote harmony and cooperation. It should show how honesty and reliability strengthen relationships and help communities function.
**Honesty prompt**—The story should describe how characters win trust, respect, or cooperation by adhering to principles of honesty. It should portray honesty as playing a positive role in personal development, interpersonal relationships, or the maintenance of social order.
**Altruism prompt**—The story should describe how characters help, protect, or share resources with others in ways that improve recipients' lives and foster social warmth and harmony. The plot should show how altruistic behavior strengthens social cohesion, treats helping the vulnerable as a shared value, and builds the helper's social reputation.
**Effort prompt**—The story should describe how characters overcome difficulties through struggle, persistence, and hard work, ultimately achieving success, earning respect, or reaching important goals. It should portray effort as a decisive driver of personal growth and social accomplishment.
**Patience prompt**—The story should describe how characters rely on patience to overcome difficulties by waiting for the right moment or sustaining long-term effort, and ultimately succeed. It should emphasize patience as valuable in learning, careers, social relationships, or life planning.
**Buddhism prompt**—The story should mention Buddhist-related content, including teachings, practices, or interactions with Buddhist figures.
**Ancestor worship prompt**—The story should center on reverence for ancestors, showing ancestors as symbols of wisdom, virtue, and spiritual strength that shape descendants' beliefs, moral choices, and life paths. It should portray ancestor worship as central to family continuity and spiritual support, with characters drawing strength from ancestral teachings, dreams, rituals, ancestral halls, or genealogical stories when facing hardship or moral dilemmas.
**Clan prompt**—The story should describe strict rules and hierarchical order within a lineage or clan, emphasizing members' obedience to clan authority. It should show characters following elders' directives, defending clan honor, or fulfilling family responsibilities, even when doing so requires personal sacrifice or creates moral conflict.
**Gift exchange prompt**—The story should center on giving and reciprocating gifts, showing gift exchange as an important way to build relationships, express respect, or fulfill obligations. It should illustrate how reciprocal gifting reinforces social ties, maintains harmony, or signals social standing.
**Benevolent emperor prompt**—The story should depict how a wise and virtuous emperor governs with benevolence, justice, and sound judgment, bringing social stability, economic prosperity, and the well-being of ordinary people. It should show how the emperor's decisions win popular support and sustain long-term order, portraying him as an ideal sovereign.

### *C.2 Synthetic Archetypes*

**Trust SA**—

In a remote village lived an elderly man everyone called Uncle Li. He was honest and always ready to help, so the villagers trusted him deeply. Whenever someone ran into trouble, Uncle Li was the first to lend a hand. The village also had a young man named Xiaoming. He was quiet and kept to himself, and the villagers kept their distance from him.

One day something strange happened. The village grain stores began to disappear, and everyone panicked. The village head called a meeting. Many villagers accused Xiaoming, arguing that he was an outsider and must have taken the grain. Xiaoming felt wronged and tried to explain, but no one wanted to listen.

Uncle Li watched this unfold and decided to speak up. He said, "Xiaoming is a good kid. I believe he would not do something like this." Because the villagers trusted Uncle Li, the room quieted down. Uncle Li suggested that everyone look for clues together, starting with Xiaoming's home.

The villagers followed Uncle Li to Xiaoming's house and, in a corner, they did find some grain. Xiaoming explained that he had bought it outside the village and had set it aside for himself and his mother. Uncle Li's willingness to trust him made everyone realize how unfair their prejudice had been.

After that incident, villagers began to approach Xiaoming more warmly, and little by little he became part of village life. Uncle Li's trust and integrity dissolved suspicion and helped restore cooperation. The villagers came to understand that trust is the foundation of good relationships, and that honesty and reliability strengthen a community's resilience. From then on, the village grew more united, helped one another more often, and lived with greater peace.

**Honesty SA**—

In a remote village lived a young man named Li Cheng. He was known for being truthful, and the villagers trusted him. One day an outsider arrived. He called himself a merchant, but his real aim was to cheat villagers out of money. He set up a stall at the village entrance and claimed he sold a "miracle potion" that could cure any illness. It was cheap, and many villagers lined up to buy it.

Li Cheng felt uneasy. He knew the potion was useless and might even be harmful. He decided to tell the villagers the truth. He walked up to the stall, cleared his throat, and called out, "Everyone, do not believe him. This potion does nothing."

The outsider snapped back, "Who are you to say it is useless. Plenty of people have bought it."

Li Cheng stood his ground. "I grew up here. I have seen too many scams. Your potion will hurt people." Hearing Li Cheng, villagers began to doubt the outsider.

Angry, the outsider proposed a contest. "If you do not believe me, let me compete with Li Cheng and we will see whose potion works."

Li Cheng knew it was a trap, but he agreed to protect the villagers. The outsider produced his potion, while Li Cheng brought a real herbal remedy from the village herb shop. They agreed to hold the contest in the village square.

On the day of the contest, villagers gathered to watch. The outsider drank his potion and then put on a show, pretending he had been gravely ill. Nothing changed, and his performance convinced no one. Li Cheng drank the herbal remedy and appeared refreshed and clear-headed. The crowd murmured, then turned on the outsider. In the end, the outsider packed up his stall and left the village in disgrace.

Li Cheng's commitment to honesty earned the villagers' respect and trust. People praised him for protecting the community, and villagers became more cautious and more attentive to truth over easy promises.

**Altruism SA—**

In a remote village lived an elderly man everyone called Uncle Li. Though old, he was kind-hearted and often helped younger villagers. There was a child named Xiaohua whose family was poor. Xiaohua's parents were often ill, and life was difficult. Uncle Li noticed that after school Xiaohua still had to work in the fields, and he felt sorry for him.

Uncle Li decided to help. Each day he went to the fields early and finished some of the work so Xiaohua could study. Xiaohua was grateful and sometimes brought vegetables from home as thanks. Uncle Li always smiled and said, "Helping you is the right thing to do. No need for gifts."

Over time, Xiaohua's grades improved. With Uncle Li's support, he was admitted to the county middle school. Villagers felt proud of the boy and praised Uncle Li's generosity. Xiaohua promised himself that he would help others when he grew up.

But a young man named Wang Er mocked Uncle Li. He said Uncle Li was wasting time and told Xiaohua, "If you rely on Uncle Li, how will you ever stand on your own." Xiaohua felt uneasy, yet his gratitude did not change.

Then a flood hit the village. Many homes were submerged and people ran for safety. Uncle Li saw that Xiaohua's house was in danger and rushed to help. In the chaos, Wang Er also became trapped. Uncle Li jumped into the water without hesitation. He rescued Xiaohua first, then turned back and pulled Wang Er to safety as well. Wang Er wept and finally understood the value of Uncle Li's kindness.

After the flood, villagers worked together to rebuild. Uncle Li's actions spread through the village, and people began to imitate his spirit of helping others. The village became more cooperative and supportive. Xiaohua later became a teacher and continued the tradition by helping children in need.

**Effort SA—**

In a remote village lived a young man named Xiaoming. Since childhood he dreamed of becoming a skilled carpenter, but villagers said he lacked talent and often laughed at him. Xiaoming did not give up. Every day he practiced under an old tree outside the village, sawing wood with a worn-out saw. At first his cuts were crooked and his work looked clumsy, but he never lost heart.

One day a famous master carpenter visited the village. People surrounded him, eager to learn. Xiaoming gathered courage, stepped forward, and asked for guidance. The master glanced at Xiaoming's work and shook his head. "This is rough. It is not worth discussing." Xiaoming felt hurt, but he practiced even harder.

Day after day his skill improved. Villagers began to notice. Then the village announced a woodworking contest. The winner would receive a prized tool. Xiaoming entered even though the competition would be intense.

On the day of the contest, villagers crowded in, expecting the master's entry to win. The master displayed a polished, intricate piece and drew loud applause. When it was Xiaoming's turn, he took a deep breath and presented a chair he had built with care. It was not flashy, but it felt warm and sturdy, with a simple, honest beauty.

After careful evaluation, the judges announced Xiaoming as the winner. People congratulated him, and even those who once mocked him began to ask him for advice. Xiaoming proved through persistence that effort can overcome obstacles and earn respect. In time he became the village's most sought-after carpenter and opened a small workshop school to teach other young people to pursue their goals.

**Patience SA—**

In a remote village lived a young carpenter named Li Ming. He loved woodworking and dreamed of making a beautiful wooden chair for the village elders. But his skill was limited. The chairs he made were often crooked or unstable, and villagers laughed at him as a beginner.

Li Ming did not quit. He searched the woods for suitable timber and patiently observed how different trees grew, learning which materials would work best. He often sat quietly under a tree for hours, thinking about how to improve his craft. Village elders, seeing his persistence, began to take him more seriously.

One day an experienced craftsman named Master Wang came to the village. Hearing that Li Ming wanted to learn, Master Wang offered to teach him. Li Ming accepted eagerly. Master Wang showed him how to use tools, how to smooth wood, and how to design a chair that would hold weight. Whenever Li Ming made mistakes, Master Wang corrected him patiently. Li Ming listened carefully and worked without slackening.

After months of practice, Li Ming's craftsmanship improved greatly. He attempted more complex designs. Each chair required time and effort, but he kept going. Villagers noticed the steady improvement, and some began to place orders.

Finally, on a bright day, Li Ming completed the chair he had long imagined. It was graceful in form, stable in structure, and comfortable to sit on. Village elders tried it and praised him warmly. Li Ming gained respect and learned that patience and steady effort over time are the keys to real skill. He continued his work, aiming to create more pieces that would serve the village.

**Buddhism SA—**

In a remote village lived a young man named Afu. He was good-natured, but he loved small gains and often used little tricks at the market to earn extra coins. Villagers knew his schemes and laughed behind his back.

One day Afu met an old monk with white hair. The monk seemed detached from worldly concerns. Afu decided to tease him. He asked, "Master, how can I get rich and make a fortune."

The monk smiled and replied, "If you want wealth, first set down the greed in your heart." Afu scoffed and walked away.

Days later, Afu saw a wealthy merchant searching for a lost gold chain. Afu found the chain and returned it, expecting a reward. The merchant thanked him and paid him. Afu felt pleased, but the old monk appeared again and said, "You are still chasing small gains."

Afu defended himself, saying he had done a good deed. The monk shook his head and replied, "A deed can look good from the outside and still be driven by greed. Real wealth is a peaceful mind." Afu was shaken. He began to reflect on his choices, gave up his small tricks, and started helping neighbors with sincere effort. Slowly he found a calmer kind of contentment, and villagers began to respect him.

**Ancestor worship SA—**

In a remote village lived a young man named Li Ming. His family had farmed for generations, and elders taught him to respect the ancestors and carry forward the family's agricultural know-how. Every spring, the village held a grand ancestral rite, and Li Ming offered incense with deep respect and prayed for a good harvest.

Trouble struck when the nearby river dried up and crops began to wither. Villagers worried and argued over what to do. Li Ming was anxious. Remembering ancestral teachings, he went to the ancestral hall to seek guidance.

That night Li Ming had a dream. He saw his grandfather standing in golden rice fields, smiling. His grandfather said, "Water is the source of life. Find the water source and you can save our fields." Li Ming awoke startled and decided to follow the dream.

He searched the area around the village for days. Finally, at the foot of a mountain he found a hidden stream. He returned to tell the villagers. Many doubted him, but Li Ming insisted. Together they dug a channel and diverted the stream into the fields.

As water flowed in, the crops revived. Villagers cheered and thanked Li Ming for his resolve. At the next rite, Li Ming prepared the best offerings and knelt in the ancestral hall. "Thank you for watching over us. We found water and saved the harvest. We will continue to inherit your wisdom and guard this land."

The story was passed down as a reminder that ancestral tradition can guide moral choices and sustain families through hardship.

**Clan SA—**

In a remote village lived the Li clan. The clan followed strict rules, and elders' orders carried the weight of law. The clan patriarch, Old Master Li, regularly convened meetings to stress honor and responsibility.

At one meeting, Old Master Li announced that the village would soon hold the annual harvest rite. As the largest household, the Li clan had to take primary responsibility for organizing the ceremony. He said sternly, "This is reverence for our ancestors and also our duty. Anyone who refuses is disloyal to the clan."

The patriarch's third son, Li Ming, felt conflicted. He had received an offer for a well-paid job in the city, a chance for a bright future. Yet he knew clan honor mattered above personal plans. He swallowed his disappointment and accepted the assignment.

On the eve of the rite, Li Ming and his family worked nonstop preparing offerings and arranging the ritual space. He did not voice his private regret. On the day itself, Li Ming wore formal traditional clothing and stood before the altar, praying for his ancestors. He kept a composed smile and did his best to show loyalty to the clan.

Afterward, villagers praised the Li clan's ceremony. Old Master Li commended Li Ming, calling him a pride of the family. Months later, when the city opportunity returned, Li Ming declined and stayed in the village to support the clan. He became a steady pillar of the family, and his story spread as an example that clan responsibility and honor can outweigh personal gain.

**Gift exchange SA—**

In a small village lived a warm and gentle old woman everyone called Grandma Li. Though elderly and not in strong health, she often helped others, and villagers respected her. Each New Year, villagers brought gifts to Grandma Li to express gratitude. Grandma Li accepted with a smile and felt comforted.

One year a young man named Xiaoming returned from working away. He had heard Grandma Li's reputation and decided to bring her a gift. He bought a basket of fresh fruit at the market and carried it to her home.

Grandma Li opened the door and beamed. "Xiaoming, what a good child. I love these fruits." Xiaoming felt pleased, thinking his gift had brought joy.

A few days later, Grandma Li decided to return the kindness. She made a pot of fragrant red bean soup, one of Xiaoming's favorites, and brought it to his home herself. Xiaoming was moved and told her he had only done what anyone should do.

Grandma Li shook her head and said, "In our village we value reciprocity. You brought fruit, so I bring soup. That is how relationships grow stronger."

After that, Xiaoming and Grandma Li became close friends, and others followed their example. Villagers came to see gifts as more than material exchange. Gifts were a way to express respect and maintain social ties, and reciprocal giving helped keep village relationships warm and stable.

**Benevolent emperor SA—**

> On the ancient land of China there was an emperor named Mingde. He governed with benevolence and was loved by the people. Mingde often traveled in disguise to learn about ordinary life. One day he came to a small village and found villagers distressed by a dried-up river.
>
> Mingde gathered villagers and asked what had happened. They explained that without water their fields were failing and hardship was growing. Mingde decided to restore the river. He issued orders at once, sending engineers and soldiers to dig the riverbed and reopen channels to the source.
>
> After months of hard work, the river returned to life. Clear water flowed again and nourished the land. Villagers celebrated and came to thank Mingde. Mingde responded modestly, saying that collective effort made the repair possible.
>
> Mingde believed governing required fairness and practical wisdom. He established a petition hall, encouraging people to submit letters and speak about local needs. When suggestions were sensible, he listened and acted. Under his rule, the state became more stable, agriculture recovered, and people lived with greater security. Stories about Mingde were told for generations as an example of benevolent rule and long-run order.

### *C.3 Validating the Synthetic Archetypes*

This subsection validates that the Synthetic Archetype (SA) narratives used in the SA–BERT similarity construction correspond to the intended themes. I implement a separate "theme-check" procedure that evaluates whether a given story provides textual evidence for a given theme, using a discrete 0–4 rubric.

**Theme-check rubric and prompt.** For each story–theme pair, I use an independent LLM evaluator to score the story's thematic fit using a strict rubric ranging from 0 (no textual evidence) to 4 (theme is the core causal backbone of the story). The instructions explicitly require that the score be based only on explicit plot evidence, discourage inference beyond the text, and recommend caution in assigning 4; the evaluator returns a single JSON object containing only the integer score. The evaluator is provided with the relevant theme definition (e.g., trust, honesty, altruism, effort, patience, Buddhism, ancestor worship, clan, gift exchange, benevolent emperor).

**Validation design and summary.** Let $s$ index an SA narrative and let $k$ index a theme definition. The theme-check exercise produces a score $Score_{s,k} \in \{0, 1, 2, 3, 4\}$ for every pair $(s, k)$. I summarize the resulting score matrix by comparing, for each theme, the score of the target archetype on its own theme (the diagonal entry) to the mean score of all other archetypes on that same theme (the off-diagonal mean). Figure A.5 reports this comparison.

**Implementation and summary.** I apply the theme-check rubric to each SA narrative against the full set of theme definitions, and then summarize the resulting matrix by averaging the scores across SA versions (when multiple versions exist for a theme). Figure A.5 summarizes the results by comparing, for each theme, the score of the target archetype on its own theme to the mean score of all other archetypes on that same theme.

**Interpretation.** Figure A.5 shows that, for each theme, the archetype written for that theme receives a much higher theme-check score than the average score received by archetypes written for other themes. The remaining overlap is concentrated among conceptually adjacent themes (for example, trust with honesty; effort with patience; and clan with ancestor worship), consistent with the fact that folktale narratives often bundle related moral content.

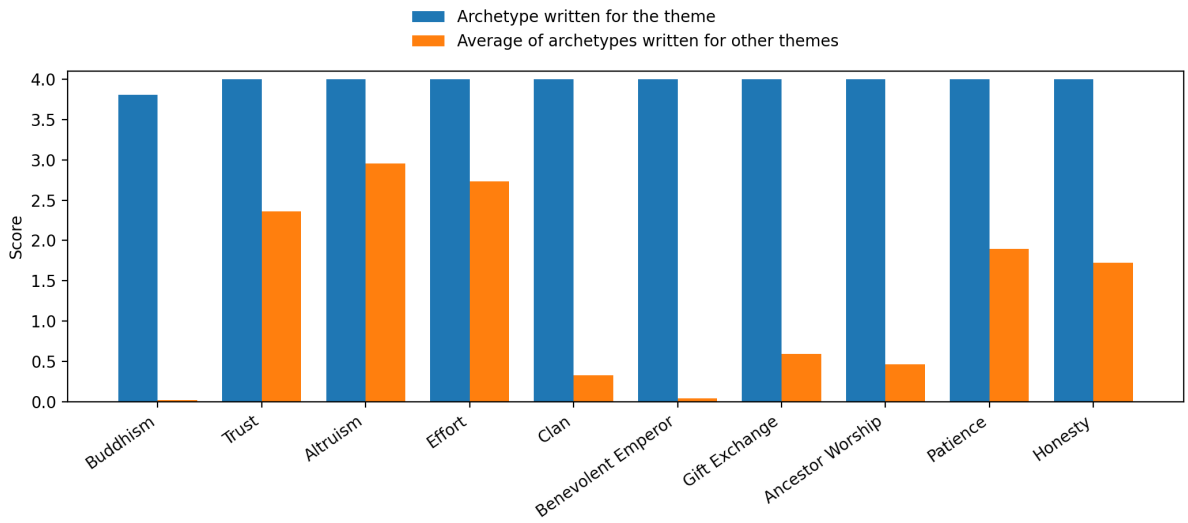
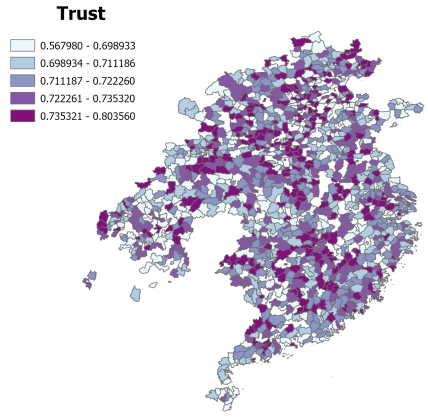


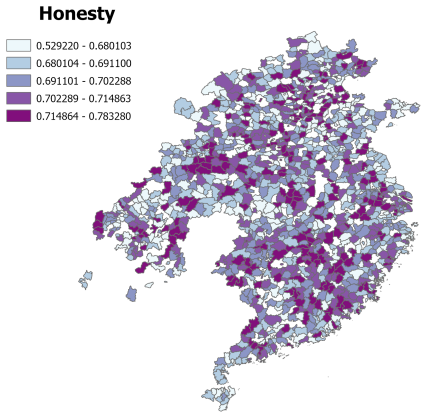

Figure A.5: Theme-check validation of Synthetic Archetypes

Note: The figure summarizes the theme-check score matrix (0–4) obtained by evaluating each Synthetic Archetype narrative against each theme definition. For each theme on the horizontal axis, the blue bar reports the theme-check score for the Synthetic Archetype written for that theme (the diagonal entry). The orange bar reports the mean theme-check score of all other Synthetic Archetypes when evaluated against that same theme (the off-diagonal mean). Higher values indicate stronger textual evidence that the narrative matches the theme; a score of 4 is reserved for cases where the theme is necessary for the main causal chain of the story.

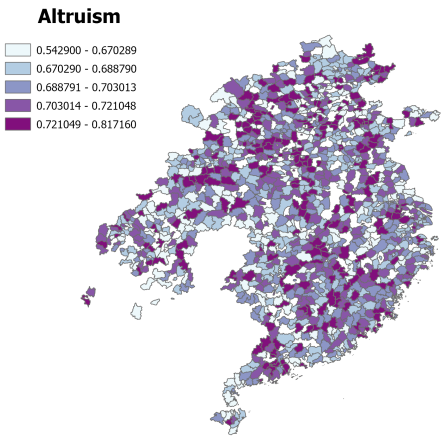
## *C.4 Maps*



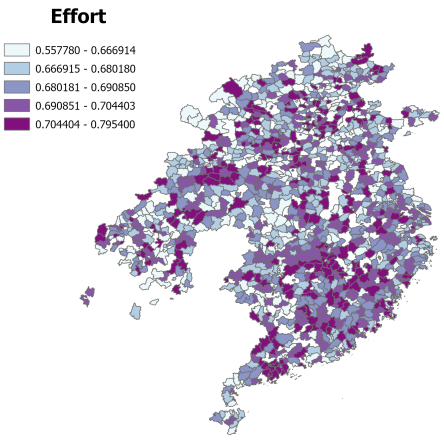

(a) Trust



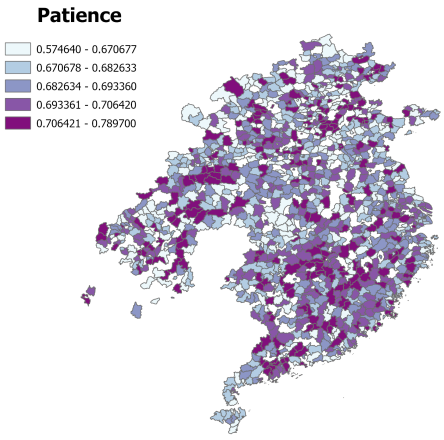

(b) Honesty



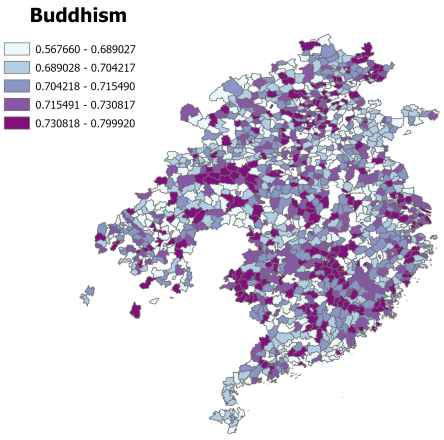

(c) Altruism



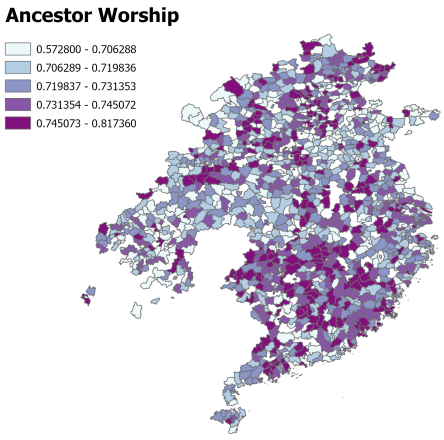

(d) Effort



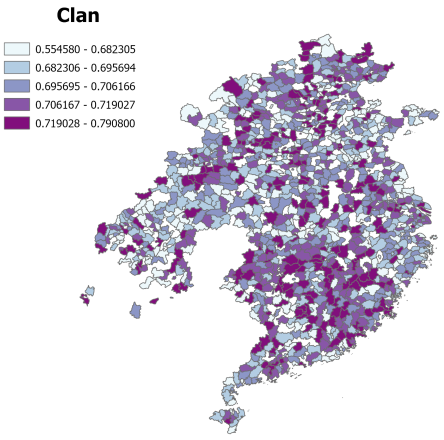

(e) Patience



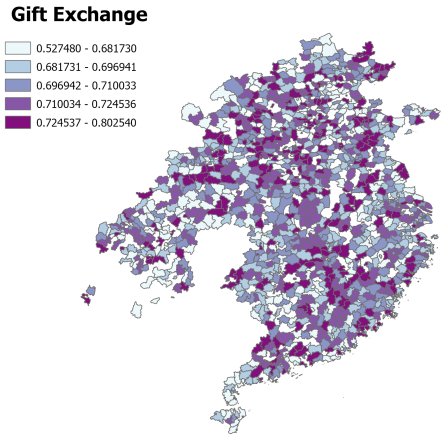

(f) Buddhism



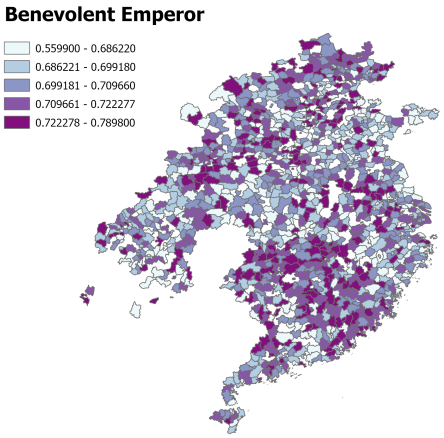

(g) Ancestor Worship




(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.

# D Folklore and Modern Attitudes

## *D.1 Patience-valuing and the "save first" item*

Table IV shows a discrepancy between columns 9 and 10 for the saving-preference outcome. In column 9, patience-valuing is positively associated with the "save first" response, while in column 10 the estimate declines and becomes statistically indistinguishable from zero once individual controls are included. To clarify what drives this attenuation, Table A.1 adds individual covariates in steps.

Column 1 reports the baseline specification with the same county-level control blocks used in Table IV. Column 2 adds gender, leaving the patience coefficient essentially unchanged. Column 3 adds age-group fixed effects, and the coefficient declines in magnitude. This suggests that part of the baseline association reflects cross-county differences in the age composition of respondents, since saving preferences and constraints vary over the life cycle and cohort. In China, the age profile of saving is unusual and older households often exhibit relatively high saving rates, so compositional differences in age can matter for the "save first" item (Chamon and Prasad, 2010).

Column 4 adds educational-attainment fixed effects on top of age groups. The coefficient declines further and loses significance entirely. This is consistent with educational attainment absorbing substantial variation in saving attitudes that is correlated with patience-valuing. In related work, long-term orientation is strongly associated with educational performance (Figlio et al., 2019).

## *D.2 Robustness*

Table A.1: Patience and Saving Preference, Individual Controls

| | "Save first" | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Patience-valuing | 1.94** | 1.89** | 1.48* | 0.93 |
| | (0.84) | (0.83) | (0.83) | (0.76) |
| Female | | 0.13*** | 0.14*** | 0.14*** |
| | | (0.030) | (0.029) | (0.030) |
| Age-group fixed effects | | | ✓ | ✓ |
| Education fixed effects | | | | ✓ |
| County-level controls | ✓ | ✓ | ✓ | ✓ |
| Mean of dep. var. | 4.07 | 4.07 | 4.07 | 4.07 |
| $R^2$ | 0.030 | 0.034 | 0.052 | 0.068 |
| Observations | 5,469 | 5,469 | 5,468 | 5,468 |

Note. The dependent variable is the CGSS item translated as "If I have extra money, I first consider saving it," coded so higher values indicate a stronger saving-first orientation. Patience-valuing is the county-level folktale-based measure for patience, constructed from tale-level cosine similarities to patience archetype tales, averaged within county, and averaged over five archetype draws. County-level controls are the full set of style, contemporary, historical, and geographic controls used in Table IV. Standard errors clustered at the county level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: 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 | 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 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Historical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Geographical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $R^2$ / 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 (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.3: 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 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Geographical Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $R^2$ / 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# Appendix References

Chamon, Marcos D., and Eswar S. Prasad. 2010. "Why Are Saving Rates of Urban Households in China Rising?" *American Economic Journal: Macroeconomics* 2 (1): 93–130. https://www.aeaweb.org/articles?id=10.1257/mac.2.1.93.

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.