Learning is Caring: Soil Heterogeneity, Social Learning and the Formation of Close-knit Communities*

Itzchak Tzachi Raz†

The Hebrew University of Jerusalem

December 2020

[Latest version available here]

Abstract

This paper studies the impact of social learning on the formation of close-knit communities. It provides empirical support to the hypothesis, put forth by the historian Fred Shannon in 1945, that local soil heterogeneity limited the ability of American farmers to learn from the experience of their neighbors, and that this contributed to their "traditional individualism." Consistent with this hypothesis, I establish that historically, U.S. counties with a higher degree of soil heterogeneity displayed weaker communal ties. I provide causal evidence on the formation of this pattern in a Difference-in-Differences framework, documenting a reduction in the strength of farmers' communal ties following migration to a soil-heterogeneous county, relative to farmers that moved to a soil-homogeneous county. Using the same design, I also show that soil heterogeneity did not affect the social ties of non-farmers. The impact of soil heterogeneity is long-lasting, still affecting culture today. These findings suggest that, while understudied, social learning is an important determinant of culture.

Keywords: Culture, Individualism, Community, Social Learning, Agriculture, Persistence

JEL Classification: N51, N52, N91, N92, N31, N32, Z10, Z13, O13, D83, D70

*I am grateful to my PhD advisors Nathan Nunn, Alberto Alesina, Melissa Dell, and Claudia Goldin for their support and guidance. I also benefited from conversations with Omar Barbiero, Anne Sofie Beck-Knudsen, Augustin Bergeron, Moya Chin, Benjamin Enke, Martin Fiszbein, Edward Glaeser, Nir Hak, Alejandro Lagomarsino and Jonathan Roth. This project also benefited from comments by conferences and seminars participants at Harvard University, Hebrew University, Ben-Gurion University, and IEHA 2020 meeting. I thank Jesse Graham and Jonathan Haidt for sharing the MFQ data. I acknowledge financial support from the Economic History Association, the Center for American Political Studies at Harvard University, and the Institute of Humane Studies. All remaining errors are my own.

†Department of Economics and PPE, The Hebrew University of Jerusalem, Jerusalem 91905, Israel.
Email: iraz@mail.huji.ac.il; Website: www.tzachiraz.com

1 Introduction

Social learning has been the focus of a large literature in economics, dating back to at least Griliches (1957). The majority of the existing research has emphasized the "learning" component in "social learning," while the "social" component stood for the form of learning, and often, for the underlying structure of social ties that might affect its efficiency. This paper emphasizes the "social" component, and asks– what happens to the strength of communal ties when meaningful and beneficial learning takes place through relationships with others?

An ideal experiment to answer this question would randomize individuals to environments that enable varying degrees of social learning, allow them to form communities, and study the short- and long-run impact on culture. Such an experiment is of course not feasible, but history provides a variation that comes very close. During the settlement of the United States, millions of individuals migrated to new environments with unknown characteristics, hoping to establish their own farm. Their resources were limited, and their survival on the new land depended on the ability to produce sufficient yield to support themselves and their families. This required them to quickly discover the optimal location-specific farming practices. One possible strategy farmers could follow was "learning by doing," i.e. an individual trial and error. Another and potentially more efficient strategy was to engage in social learning and to build on the experience of their neighbors. However, substantial heterogeneity of soil in their area implied that the optimal farming practices were highly local, which limited the effectiveness of social learning. According to the historian Fred Shannon (1945), farmers' inability to rely on learning from others fostered their "traditional individualism."

I provide the first empirical evidence supporting Shannon's *"Social Learning Hypothesis."* I use detailed spatial soil data to construct a novel county-level measure of local soil heterogeneity. Then, using this measure, I establish a negative historical association between soil heterogeneity and a close-knit social structure. I also provide suggestive evidence supporting limited social learning as a probable channel. I then proceed to document and discuss the persistence of this association. Finally, I provide causal evidence on its formation in a Difference-in-Difference framework, documenting the impact of soil heterogeneity on the culture of farmers that migrated within the U.S.

To measure local soil heterogeneity that is most relevant for the effectiveness of farmers' social learning I rely on soil taxonomy. The taxonomy is designed to facilitate predictions of agricultural output under different systems of management, implying that two plots of land with different optimal farming practices are differentiated into different taxonomic classifications. I use detailed soil data from the Digital General Soil Map of the United States (STATSGO2) (Soil Survey Staff, 2017b) to construct a county-level *"Soil Heterogeneity Index"* (SHI) for counties in the contiguous U.S. The

SHI captures the average degree of dissimilarity of soil across neighboring locations in the county. It ranges from 0 to 1 and equals the average probability that a given location in the county has different soil properties than a randomly selected location close-by.

A fundamental challenge in historical research on the strength of communal ties is the lack of data. I overcome this challenge by proxying close-knit communities with the centrality of communal identity in individuals' self-definition, and inferring the later from children naming patterns. This approach builds on two strings in the literature. First, it builds on the seminal contributions of Hofstede et al. (2010), Triandis (1995), and Markus and Kitayama (1991), which highlight that members of close-knit cultures tend to define themselves with reference to groups. The centrality of relationships with in-group members in individuals' self-definition is considered the fundamental difference between close-knit and loose-knit social structures. Triandis et al. (1990) provide empirical support for the use of the importance of social groups in individuals' self-identity as a proxy for close-knit social networks.

Second, it builds on a rich literature that uses naming patterns to identify cultural tendencies in historical data, based on the assumption that first names carry information regarding parental culture and social identity. In particular, naming patterns have been used in the economics literature to measure social identification with different groups, such as race (Fryer and Levitt, 2004), ethnicity (Fouka, 2019), a nation (Abramitzky et al., 2020; Russo, 2019), and socioeconomic status (Olivetti and Paserman, 2015). To infer the communal identity content in children's first names, I follow Fryer and Levitt (2004) and construct a *"Local Name Index"* (LNI) using children's first names in the full count census data between 1850-1940. The LNI captures the probability that a name is given to a "local" (i.e., from the same county or state) child relative to a child in different locations in the U.S. A high value implies a distinctively local name. I validate the LNI as a meaningful measure of close-knit social structure using contemporary data from Enke (2020).

Using those two novel measures I document the robust negative historical association between soil heterogeneity and close-knit communities. I find that communities in counties with a high local soil heterogeneity exhibit a loose-knit social structure. That is, parents living in high local soil heterogeneity counties have a lower tendency to choose names that signal the local identity. In my preferred specification, an increase in local soil heterogeneity from a complete homogeneity (SHI = 0) to a complete heterogeneity (SHI = 1) is associated with about 2.49 ($p-value < 0.001$) points decrease in the county's average LNI. This association is robust to many variations in measurement and specification, including alternative definitions of the LNI, alternative definitions of the SHI, and using alternative ways to account for spatial auto-correlation.

I also document a similar impact of soil heterogeneity on two alternative measures of close-knit social structure: religious diversity and the strength of family ties. I construct a religious diversity index using county-level data on the number of members in religious institutions by denomination. The index captures the degree to which multiple cultural-religious identities exist within a community. I construct a measure of the strength of family ties using census data on family structure and the choice of living arrangements. I find that an increase in local soil heterogeneity is associated with a higher religious diversity and weaker family ties. Those findings suggest that the impact of soil heterogeneity on culture and personality may have originated with communal ties, but eventually extended to other forms of social relationships and in-groups.

I then proceed to provide suggestive evidence supporting a social learning hypothesis interpretation of the those patterns. First, I show that the relationship between soil heterogeneity and close-knit communities remains mostly unaffected when I additionally control for the share of farmers. This suggests that the reduced-form association with culture is not explained by an impact on the development of non-agricultural industries or urbanization more generally (Fiszbein, 2019). Second, I show that soil heterogeneity does not matter for culture if there are no farmers in the county, and that it matters more the more farmers there are. Although not causally identified, those results are suggestive of an association between soil heterogeneity and culture that operates through an impact on farmers' attributes and social ties. I also document a positive relationship between local soil heterogeneity and actual agricultural diversity, as expected if indeed soil heterogeneity implied that optimal farming practices were highly local and limited the effectiveness of farmers' social learning. Finally, I find a negative association between local soil heterogeneity and the rate of fertilizers adoption, which is consistent with soil heterogeneity limiting farmer's social learning.

Having established a reduced-form historical association between local soil heterogeneity and close-knit communities, I continued to study its persistence. First, I establish that the impact of soil heterogeneity weakened over time, as the impact of an increase from a complete soil homogeneity to a complete soil heterogeneity on children's LNI drops from -4.98 ($p-value < 0.001$) in 1850 to -1.00 ($p-value = 0.218$) in 1940. Second, using survey data on individual's moral values (Graham et al., 2011), I document a negative association between soil heterogeneity and communal morality in the long-run. An increase from a complete soil homogeneity to a complete soil heterogeneity in the county of residence is associated with about 0.05 standard deviation drop in respondents' communal moral values, with no impact on individualizing moral values.

My preferred interpretation for this pattern focuses on cultural persistence, rather than a continuing impact of soil heterogeneity on farmers' social learning. The massive migration of farmers has long ended, the share of the population engaged in agriculture declined, and farmers' access to precise information on soil management had substantially grown. Those facts might explain the weakening of the association between the SHI and the LNI over time. While social learning may continue to shape local culture in general, farmers' limited ability to engage in social learning due to soil heterogeneity is unlikely to be a central force that continues to actively shape local culture today. Yet soil heterogeneity shaped the nature of social relationships at "critical juncture" in history– when farmers first arrived at new locations, new communities were formed, norms and ways of living were determined, and local institutions were established. Under those circumstances, a long-lasting impact on culture seems quite plausible.

Finally, I provide causal evidence on the formation of the association between soil heterogeneity and close-knit communities. I focus on the nineteenth century and exploit within-family variation in naming patterns in a Difference-in-Differences framework across families that migrated within the U.S. First, I use children's year and state of birth to identify families that moved across states. Then, using an LNI measure in which the state of birth is defined as "local," I study the naming pattern of children in families that migrated to counties with varying degrees of soil heterogeneity, born before and after their family had moved.

I document a decrease in communal identification among farmers that moved to soil-heterogeneous counties, relative to farmers that moved to soil-homogeneous counties, with no impact across non-farmers migrating to the same locations. The LNI of children born to farmers after they had moved to a county with a complete soil heterogeneity decreased by 3.25 ($p-value < 0.001$) points, relative to the change in the LNI of farmers' children born after a move to a soil-homogeneous county. Importantly, there are no differences in the levels of children's LNI born before the move across families that migrated to different soil heterogeneity environments. This lack of pre-trends suggests that there was no selective migration on prior levels of communal identification. When I study non-farmers' naming patterns I find a similar lack of pre-trend, but also a lack of differences after the move. The causal impact of soil heterogeneity on the culture of farmers along with a null impact on that of non-farmer provides strong support to the social learning hypothesis. Similar to the correlational evidence above, this pattern substantially reduces the likelihood that the relationship between soil heterogeneity and close-knit communities is explained by channels different than an impact on farmers' attributes and social ties.

Related Literature. Considerable research across the social sciences focused on the cultural, psychological, moral, and political differences between close-knit versus loose-knit social structures. Often referred to as the most fundamental difference across cultures, this cleavage has been referenced to in the literature by different terms, such as *"Individualism"* versus *"Collectivism"* (Hofstede et al., 2010; Triandis, 1995), *"Independence"* versus *"Interdependence"* (Markus and Kitayama, 1991), and *"Autonomy"* versus *"Embededdness"* (Schwartz, 1994). While the vast majority of the literature had focused on documenting the cleavage and studying its effects, a smaller literature explored its historical roots.

According to the influential evolutionary approach to culture (Boyd and Richerson, 1988), the two cultural syndromes may have been chosen by different societies because each of them may be a successful strategy in a particular environment. Past work provided empirical evidence supporting an ample of different hypotheses regarding certain aspects of the environment that may be conducive to either a close-knit or a loose-knit social structure, including the in-group cooperation hypothesis (Talhelm et al., 2014; Buggle, 2018; Ang, 2019), the modernization hypothesis (Greenfield, 2009), the pathogen prevalence hypothesis (Fincher et al., 2008), and the voluntary-settlement hypothesis (Turner, 1921; Kitayama et al., 2006; Varnum and Kitayama, 2011; Bazzi et al., 2020; Beck-Knudsen, 2019) or the more general residential mobility hypothesis (Oishi et al., 2007, 2009). This paper contributes to this interdisciplinary literature by advancing the *"Social Learning Hypothesis"* as another potential cause. Although this hypothesis was put forth by the historian Fred Shannon 75 years ago, it received very little attention in the literature that followed. This is the first paper to provide causal evidence supporting it.

Apart from Shannon's historical writing, the only other existing research I am aware of that points to social versus individual learning as a possible explanation for close-knit versus loose-knit social structure is Chang et al. (2011), which argues that East–West cultural differences resulted from differential use of social versus individual learning in response to differences in the degree of stability of the environment over time. However this study focus on environmental variability over time, and not across space,[1] and does not provide any causal evidence or even direct correlational evidence. Other studies in social psychology have also associated social learning with a close-knit social structure, but they seem to think of causality as running from culture to social learning, and not the other way around (Yaveroglu and Donthu, 2002; Mesoudi et al., 2015).

The findings of this paper speak more generally to two areas of research- a literature on the historical roots of culture, cultural persistence and cultural change (e.g. Bisin and Verdier, 2001; Nunn and Wantchekon, 2011; Alesina et al., 2013; Galor and Özak, 2016; Abramitzky et al., 2020; Giuliano and Nunn, 2020), and the broad literature about the individualism-collectivism cleavage in social psychology (Hofstede et al., 2010; Markus and Kitayama, 1991; Triandis, 1995) and in economics (Bazzi et al., 2020; Beck-Knudsen, 2019; Buggle, 2018; Gorodnichenko and Roland, 2011, 2017, 2015; Enke, 2019, 2020).

This study also adds to a growing empirical literature that has emerged in recent years focusing on the impact of farmers' settlement of the U.S. on local long-run culture and economic development (Bazzi

[1]From this prospective, this study is more closely related to the hypothesis advanced by Giuliano and Nunn (2020).

et al., 2020; Fiszbein, 2019; Raz, 2018; Mattheis and Raz, 2019; Smith, 2019). Although each line of research focuses on different aspects of the farmers' settlement, each provides supporting evidence for how formative was the period of westward expansion and how persistent are its impacts.

Finally, this paper also ties into the large literature in economics on social learning, and in particular on social learning in agriculture (Griliches, 1957; Besley and Case, 1994; Foster and Rosenzweig, 1995; Conley and Udry, 2010) and on heterogeneity and social learning (Ellison and Fudenberg, 1993; Munshi, 2004; Yamauchi, 2007). To the best of my knowledge, this is the first paper to study the effects of social learning on culture.

Outline. The paper proceeds as follows. Section 2 defines key concepts and lays out a conceptual framework. Section 3 establishes a historical association between soil heterogeneity and close-knit communities. Section 4 establishes and discuss persistence of this association, while section 5 provides causal evidence on its formation. Section 6 offers concluding remarks.

2 Definitions, Conceptual Framework and Research Hypotheses

This section defines two key term– *"Close-knit Communities"* and *"Soil Heterogenity"*, lays out a conceptual framework for thinking about the association between them, and states testable research hypotheses.

2.1 "Close-knit Communities"

Considerable research across the social sciences focused on a fundamental cleavage across cultures: in some cultures, social networks are close-knit and individuals are interdependent within their in-group. Their self-definition is tied to the in-group and their relationships with others, they tend to prioritize the goals of their in-group, pay attention to group memberships in determining their social relationships, and their behavior is largely shaped by in-group norms. In others, social networks are loose-knit. Individuals are autonomous and independent, they prioritize their personal goals over the group, and shape their behavior and relationships with others based on personal attributes.

This cultural divide is often referred to by different terms, such as *"Individualism"* versus *"Collectivism"* (Hofstede et al., 2010; Triandis, 1995), *"Independence"* versus *"Interdependence"* (Markus and Kitayama, 1991), and *"Autonomy"* versus *"Embededdness"* (Schwartz, 1994).[2,3] While those

[2] And recently in the economic literature, focusing on moral values, *"Universalism"* versus *"Communalism"* (Enke, 2020).

[3] Or to describe differences in personality attributes among individuals within cultures, *"Idiocentrism"* versus *"Allocentrism"* (Triandis, 1995, 2001).

terms may be used to highlight different aspects of social life and do not only constitute differences in terminology, by-and-large they capture the same fundamental cultural cleavage. Although this cultural cleavage is often associated with the differences between the West and East Asia, in fact differences can exist even within much smaller geographies. Significant differences were documented across regions within a country (Vandello and Cohen, 1999; Varnum and Kitayama, 2011; Talhelm et al., 2014; Bazzi et al., 2020). Moreover, individuals within the same culture may exert more or less individualist (or collectivist) personal attributes (Triandis, 2001; Graham et al., 2009; Enke, 2020).

This paper focuses on cultural differences across communities within the U.S. I adopt the terminology *"Loose-knit Communities"* versus *"Close-knit Communities"* to make this focus salient. By this I mean the extent to which communities within the U.S. are characterized by close-knit social networks, their members are interdependent, and the community constitutes an important component of individuals' self-definition.

2.2 Soil Heterogeneity

Soil is a naturally occurring mixture of minerals, organic ingredients, liquid, and gases, with a definite form, structure, and composition, resulting from a unique combination of parent material, climate, living organisms, landscape position and time. It is characterized by distinguishable layers, formally referred to as "horizons," that resulted from natural processes such as additions of materials, losses, transfers, and transformations of energy and matter. The nature of a soil depends on all of its layers, and many of its properties can not be determined from the surface alone. This implies two important facts: first, significant differences in soil properties can exist even within short distances. Second, most people are unaware of those differences (Soil Survey Staff, 1999; Soil Survey Staff, 2017a).

The multidimensional characteristics of soil are continuously varying across space. This poses a challenge for the classification of soils. Soil scientists' solution to the problem was a practical one - soil taxonomy was designed to facilitate "predictions of the consequences of specific uses of soils, commonly in terms of plant growth under specified systems of management but also in terms of engineering soil behavior after a given manipulation" (Soil Survey Staff, 1999, p. 18). Therefore, the properties that differentiate taxa are the ones that are most important for that purpose. This makes soil taxonomy particularly useful for this study, as it implies that two plots of land that are different in terms of optimal farming practices will be differentiated to different taxonomic classifications. Moreover, differentiating properties are generally not affected by cultivation or similar human activities.[4]

[4]"[...] the differentiate keep an undisturbed soil and its cultivated or otherwise human-modified equivalents in the same taxon insofar as possible. Changes produced by a single or repeated plowing that mixes the surface soil to a depth of 18 to 25 cm (7 to 10 in), for example, have the least possible effect on the placement of a soil in soil taxonomy. Truncation by

This fact is important from a research design point of view, as it significantly alleviates concerns of the endogeneity of modern-day soil classification to a particular history of human farming practices.

I use detailed geo-referenced soil data from the Digital General Soil Map of the United States (STATSGO2) (Soil Survey Staff, 2017b) to construct a county-level *"Soil Heterogeneity Index"* (SHI), which captures the average dissimilarity of soil across neighboring farmers in the county. Specifically, the index ranges from $0-1$ and measures the average probability that a given location in the county has different soil properties than a randomly selected location close-by. See Appendix D.1 for more details on the construction of the SHI.

Figure 1 plots the county-level SHI for counties in the contiguous U.S. in 2000, after partialling out state fixed effects.[5,6] A darker color implies a higher soil heterogeneity. The figure makes clear that the degree of local social heterogeneity is not uniform across the country. Importantly, there is also substantial variation in SHI within states, which is the variation that will be exploited in the preferred specifications of the empirical analysis below (sections 3 - 4).

2.3 Soil Heterogeneity, Social Learning, and Culture

American farmers that settled the frontier had to adapt to new and unfamiliar environments. Their transitional practices were often meet with failures. To succeed they had to be dynamic and innovative (Shannon, 1945; Olmstead and Rhode, 2008, 2011). However, according to the historian Fred Shannon (1945), the high degree of soil heterogeneity in the U.S. limited their ability to learn from the experience of their neighbors. Farmers often tried to follow the guidance of local agricultural society or imitate the agricultural practices of successful farmers in the area, failing to appreciate the fact that their plots different soil characteristics, and "got worse crops than before" (p. 4). Their inability to rely on social learning forced them to rely instead only on themselves, which fostered their "traditional individualism".

Unfortunately, Shannon did not discuss the explicit channel through which limit social learning strengthen farmers' "individualism." I consider two main plausible channels. First, in an environment of low soil heterogeneity, social interactions were useful since they had the potential of increasing

erosion does not change the classification of a soil until horizons or diagnostic features important to the use or identification of the soil have been lost. Consequently, insofar as possible, the diagnostic horizons and features should be those below the part of the soil affected by human activities. However, significant changes in the nature of the soil by humans cannot be ignored" (Soil Survey Staff, 1999, p. 16).

[5] I present the figure after partialling out state fixed effects since the underlying soil surveys tend to very by state. For the same reason, I also include state fixed effects in my preferred empirical specification.

[6] While the underlying soil data is assumed to be time-invariant in the empirical analysis in this paper, county boundaries changed over the years, therefore leading to different county-level SHI data for different years.

farmers' output through social learning. It thus seems plausible that in such environments farmers were more likely to invest in social relationships with their neighbors. Over time this would contribute to the development of a close-knit community and a high degree of communal interdependence.

A related channel involve an impact on personal attributes rather than incentives to investment in social relationship. In an environment where social learning is relatively more effective due to low soil heterogeneity, individuals' may eventually come to place greater value on the actions of others. This channel is similar to the hypothesis advanced by Giuliano and Nunn (2020), that the stability of the environment across generations makes it more likely that past-traditions will be optimal for the current generation. Individuals will therefore place greater value on tradition, and cultural persistence will be greater. The second channel is a horizontal version of the same logic, applied within a generation across members of the community rather than across generations, with the "interdependent-self," who is attentive to others (Markus and Kitayama, 1991), serving as an analog to the "traditionalist."[7]

Soil heterogeneity may also strengthen "traditional individualism" regardless of social learning. A third related channel focus on the heterogeneity of actions. In a relativity homogeneous soil environment, the variance in the actual farming practices was likely to be lower. Put simply, farmers in such regions were more likely to be all growing the same crops, at the same time, using the same inputs and techniques. This is likely to hold simply because the optimal practices would be more homogeneous, and would thus hold regardless of whether farmers learned them individually or socially. Over time, this could have contributed to the development of a *"tight"* culture which places a strong emphasis on norms. Tight cultures tend to be collectivist and their members tend to be interdependent (Triandis, 1995, 2001).

All three channels generate the same prediction regarding the impact of soil heterogeneity on the formation of close-knit communities with interdependent members. Since I can not directly observe social learning or investment in social relationships, I am unable to empirically disentangle those three channels.[8] It may very well be that all three channels have some merit. Yet the first two channels seem to most closely match Shannon's line of thought, and therefore, also constitute the preferred interpretation advanced by this paper. I refer to the combination of the two channels as the *"Social Learning Hypothesis"*.

[7] There is also a parallel to be drawn between the those two channels, that focus on social learning, and the hypothesis that in-groups cooperation more generally contributes to the development of a collectivist culture (Talhelm et al., 2014; Buggle, 2018). In some sense, knowledge sharing is a form of cooperation. However, even under this interpretation, it is important to note that social learning is a much more limited form of cooperation. American farmers generally did not collaborate working on the same plot, and social learning did not require them to spend a significant share of their working time with others.

[8] However in the empirical analysis below I present some suggestive evidence that specifically support social learning as the channel.

The social learning hypothesis can be reformulated in terms of *relative* soil heterogeneity to generate the following testable prediction:

Prediction 1 (Close-knit Communities)
Communities located in a higher soil heterogeneity environment are less likely to have close-knit social networks with interdependent members.

In theory, there could be other channels that may relate soil heterogeneity to close-knit communities but generate different predictions. One channel that seems particularly plausible ex-ante, is that in an environment of high soil heterogeneity there might be a greater scope for farmers to co-insure against adverse agricultural shocks. This is because different soil types imply a different level of exposure to a variety of agricultural shocks (e.g., flooding or crop-specific shocks). Note, however, that according to this channel farmers in a soil-heterogeneous location stand to benefit *more* from local cooperation and stronger social ties, relative to farmers in a soil-homogeneous location. Providing empirical support for Prediction 1, therefore, disputes the importance of this channel.

There could also be other potential channels by which soil heterogeneity might impact the formation of close-knit communities that do not involve farmers' attributes or their social engagement in particular. For example, local soil heterogeneity is likely to cause local agricultural diversity, which was found to foster industrialization and innovation (Fiszbein, 2019). More broadly, soil heterogeneity might have a direct impact on non-agricultural production, trade cost, or construction costs,[9] and therefore modernization of the economy. This, in turn, may impact the formation of close-knit communities (Greenfield, 2009). Another confounding channel may be settlers' diversity. If settlers were seeking and able to locate environments with similar properties as the one they were familiar with back home, higher local soil heterogeneity might cause a higher degree of birthplace diversity, as different soils may attract settlers from different locations. This in turn might result in a loose-knit social structure. If such confounding channels hold merit, then Prediction 1 may be validated by the data even if the social learning hypothesis is false. The following testable prediction would allow to assess the plausibility of such channels:

Prediction 2 (Soil Heterogeneity Only Impact Farmers)
Soil heterogeneity should only directly impact the attributes and communal ties of farmers.

[9] That is, after controlling for other geo-climatic characteristics.

3 The Historical Association of Soil Heterogeneity and Culture

This section studies the historical association between soil heterogeneity and culture. I discuss and validate a proxy measurement of close-knit communities in historical data, present the estimation framework, and document a robust negative association between local soil heterogeneity and close-knit communities. I also provide suggestive evidence supporting a social learning interpretation of this association, and document similar findings using other proxy measures of close-knit social structure.

3.1 Measuring the Strength of Communal Ties in Census Data

A fundamental challenge in studying the historical impact of soil heterogeneity on the strength of communal ties is the lack of systematic historical data on the later. The premise of this paper is that the degree to which local social networks are close-knit can be inferred from children naming patterns.[10] Research in social psychology documented that individual members of close-knit social networks tend to define themselves with reference to groups. The foundational study of Hofstede et al. (2010) notes that individualism versus collectivism is "reflected in whether people's self-image is defined in terms of 'I' or 'we'." The seminal work by Markus and Kitayama (1991) argues that the importance of relationships with others in individual's self-definition is the most significant difference between close-knit and loose-knit social structures.[11] The influational research by Triandis (1995) similarly identifies the definition of the self as a central defining attribute of individualism versus collectivism. Triandis et al. (1990) provides empirical support for the use of the centrality of social groups in individuals' self-identity as a proxy for close-knit social networks.[12]

My approach builds on this line of thought and uses the degree of communal identification to proxy for a close-knit social structure. I use children naming patterns to measure the extent to which parents identify with the local community. This follows a rich literature that uses first names to identify cultural tendencies in historical data. It relies on the assumption that names contain a deep cultural component since parents choose children's first names that reflects their own cultural identity, believes, and preferences. In economics, naming patterns have been used to measure parental desire to identify

[10] In section 3.5, I present two alternative measures, one that focuses on religious diversity and another that focuses on the strength of ties within the family.

[11] "The most significant differences between these two construals [independent and interdependent] is in the role that is assigned to the other in self-definition. [...] for the interdependent self, others are included *within* the boundaries of the self because relations with others in specific contexts are the defining features of the self. [...] With an independent construal of the self, others are less centrally implicated in one's current self-definition or identity" (Markus and Kitayama, 1991, p. 245-246).

[12] In Triandis et al. (1990), individuals across different cultures are asked to respond to the "I am ..." question. The percentage of responses that were linked to a social group (e.g. family, religion or location) was found to be positively correlated with collectivism.

with a racial group (Fryer and Levitt, 2004), an ethnic group (Fouka, 2019), the nation (Abramitzky et al., 2020; Russo, 2019), and socioeconomic status Olivetti and Paserman (2015), as well as to measure parental taste for uniqueness versus fitting-in (Bazzi et al., 2020; Beck-Knudsen, 2019).

To measure a first name's communal identification content, I follow Fryer and Levitt (2004) and construct a *"Local Name Index"* (LNI) using children's first names in the full count census data between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019).[13,14] The LNI is defined as

$$LNI_{first\ name,l,g,t} = 100 \times \frac{Pr(first\ name|l,g,t)}{Pr(first\ name|l,s,t) + Pr(first\ name|-l,g,t)} \tag{1}$$

where l is the geographical level defined as "local" - the contemporaneous county or state, g is the child's gender, and t is the census year. The index has an intuitive interpretation– it captures the probability that a name is given to a local child relative to a child in different locations in the U.S. It ranges from 0 to 100, where a value of 100 reflects a distinctively local name and a value of zero reflects a distinctively "outsider's" name.[15] Note that the LNI is invariant to the size of the population in different localities and to the general popularity of a given name. My main interest is in an LNI measure where "local" is defined as the county, but results are similar when local is defined as the state.

To get some intuition for the LNI measure, consider the following example. In 1940, about 0.49% of boys in the U.S. were named "Billie".[16] However, there was substantial regional variation in the popularity of the name. In Arkansas, about 2.03% of boys were given that name, while in Massachusetts only about 0.0005%. Those striking regional differences in the popularity of the name "Billie" meant that the name caries information regarding the likely state of birth. Put differently, naming their child "Billie' was a good way for parents in Arkansas to signal their group identity.[17] As a result, a "Billie" in Arkansas is assigned with a high LNI of 81.42, reflecting a name that is highly local to the state, but

[13] Fouka (2019) and Abramitzky et al. (2020) uses the same procedure to construct a "German Name Index" and a "Foreignness Index", respectively.

[14] 1850 is the first year in which first names were recorded. 1940 is the last year for which full count data is available. Data is unavailable for 1890.

[15] In constructing the LNI, I follow Bazzi et al. (2020) and restrict the baseline sample to white native-born children between the age of 0 to 10 with native-born parents to remove variation in naming patterns that are associated with demographic rather than culture. Results are robust to variation in the sample selection.

[16] That is, they appear in census records as "Billie." Their official name may very well be "William." I am agnostic to this question, as communal identification is likely, and arguably, more likely, to present itself also in the name that is used in practice in social contexts rather than only in names that appear in formal records.

[17] More generally, "Billie" is a good "Southern name".

a very low LNI of 0.11 in Massachusetts.[18,19]

Figure 2 plots the county-level LNI for the contiguous U.S. in 1940. Darker colors imply a higher LNI. Clear non-random spatial cultural patterns emerge from the map. The Northeast, the "rust belt" and the West Coast tend to have a low LNI, reflecting loose-knit social networks, while the south, the "wheat belt" and areas with a high fraction of Mormons (Utah and parts of Nevada and Idaho) tend to have a high LNI, reflecting close-knit social networks. Moreover, within states, counties that are home to large cities (e.g., Fulton County, Georgia, where Atlanta is located) tend to exhibit weaker communal ties than other counties in the state.

I validate the LNI as a measure of close-knit communities using contemporary measures of communal values from Enke (2020), which relates communal values and voting patterns. Table 1 presents the results. I find that higher LNI in 1940 (the latest year for which I have data) correlates with a higher contemporary relative importance of communal values (columns 1-2), higher vote share for Trump in 2016 (columns 3-4), and a higher 2016 Trump vote share relative to previous Republican presidential candidates (columns 5-6). Those results provide important validation for the LNI. Moreover, they suggest that differences in the degree to which communities are close-knit are persistent.

3.2 The Estimation Framework

I study the relationship between soil heterogeneity and local culture with the following estimation framework:

$$Culture_{ct} = \beta Soil\ Heterogeneity_c + \theta_{s(c)t} + X_c\Gamma + \epsilon_{ct} \tag{2}$$

where $Culture_{ct}$ is a cultural outcome of interest in county c in year t, $\theta_{s(c)t}$ is a state-by-year fixed effect, and X_c is a vector of time-invariant geo-climatic controls, which includes in the baseline specification average temperature, average precipitation, average slope, average elevation, average ab-

[18]The name "Billie" presents similar spatial patterns for girls, with an LNI of 77.09 in Arkansas and 2.43 in Massachusetts. A similar pattern also exists using the name "Billy" for boys, with an LNI of 75.74 in Arkansas and 1.09 in Massachusetts.

[19]It is important to note that the LNI is designed to capture a different variation than measures of first name commonness used in Bazzi et al. (2020) and Beck-Knudsen (2019). The LNI is destined to capture the group identity component in a name, while name commonness is design to capture a desire to fit-in versus standing out. The reason the two might generate different patterns, is that a relatively uncommon name might sometimes actually reflect a desire to signal identification with a particular group rather than a desire to stand out (i.e. and not identify with any group). As a result, even when a name commonness measure is defined over the local geography rather than the national level the LNI and a measure of first name commonness might result in different patterns. For example, only about 0.003% of boys in the U.S. were named "Waytt" in 1940, implying an uncommon name. The name was quite uncommon even in Alabama- the state with the highest share- 0.015%. Yet the name was *relatively* much more common in Alabama than outside of it, leading to a high LNI of 85.45 in Alabama, much higher than the lowest LNI of 5.30 in Pennsylvania.

solute agricultural productivity, flow accumulation, and river density. I also include a smooth control for location with a second-order polynomial in latitude and longitude, which absorbs all omitted geo-climatic characteristics that change smoothly across space. β is the coefficient of interest, representing the relationship between the degree of local soil heterogeneity and the cultural outcome of interest.

I cluster observations at arbitrary grid-cells to account for spatial auto-correlation, as proposed by Bester et al. (2011). This approach is considerably less computationally demanding in large samples compared to Conley (1999) spatial standard errors. The baseline specification uses a grid size of 100 square miles, but results are robust to using grid-cells of difference sizes, as well as other (non-arbitrary) clusters.

3.3 Main Result: Communal Identity

I find a robust negative relationship between soil heterogeneity and the strength of communal identity, proxying for close-knit communities more generally. Table 2 reports the estimates of equation 2 when the dependent variable is the county-level LNI in which "local" is defined as the county, for four different specifications. Across all specification I find that soil heterogeneity decreases the strength of communal identification, thereby providing empirical support to Prediction 1.

The specification in column 1 does not include any controls. It suggests that an increase from a complete soil homogeneity (SHI = 0) to a complete soil heterogeneity (SHI = 1) is associated with a decrease of about 4.52 ($p-value < 0.001$) points in children's LNI. Columns 2-4 add controls to address omitted variable bias concerns. In all columns, there is a large increase in R^2 relative to a modest change in β, which results in $|\delta| > 1$ and suggests that selection on unobservables is unlikely to drive this association (Oster, 2019). The specification in column 2 includes state-by-year fixed effects. The association between soil heterogeneity and children's LNI is stronger relative to column 1, as the point estimate decreases to about -5.51 ($p-value < 0.001$) points. When I add observable geo-climatic controls in column 3, the point estimate drops to -2.91 ($p-value < 0.001$). The specification in column 4, which controls both for observable and unobserved geo-climatic characteristics, is my preferred one. It suggests that children's LNI is about 2.49 ($p-value < 0.001$) points lower in counties with complete soil heterogeneity relative to counties with complete soil homogeneity.

Robustness Checks. I document high robustness of the finding in Table 2. Appendix Figure C.1 documents robustness to alternative ways to account for spatial auto-correlation in the data for inference. In all the alternatives I consider β from equation 2 is estimated with a high level of precision.

The result is also robust to alternative ways to compute the LNI in practice. Appendix Table C.1 documents that the result holds when I: (i) also include in the sample children with foreign-born parents (column 2), (ii) also include in the sample children with native-born parents of all races (column 3), (iii) include in the sample children of all races and any parental birthplace (column 4), (iv) only include in the sample names that are observed at least 100 times nationally, to filter-out variation resulting from unique spelling or transcription errors (column 5), and (v) define "local" as the state instead of the county (column 6).

Finally, the finding is robust to different ways of defining the SHI. Appendix Table C.2 reports results using alternative distances in the calculation of the SHI. The baseline SHI definition uses half of the mean county size as a benchmark for the area considered of the heterogeneity calculation (column 1),[20] however estimates are stable when I use smaller areas (columns 2-4) or larger areas (columns 5-7).

3.4 A Social Learning Story - Suggestive Evidence

The results thus far document a robust reduced-form relationship between soil heterogeneity and a strong communal identity, but they do not explicitly point to social learning and farmers' social dependence as the key channel. This section provides suggestive evidence supporting the social learning hypothesis as the explanation for the reduced-form association.

3.4.1 Soil Heterogeneity Only Matters for Agriculture

I provide suggestive evidence that soil heterogeneity's impact on culture is rooted in farmer's attitudes and social ties. The results below demonstrate that soil heterogeneity is associated with culture even holding the share of farmers fixed, that soil heterogeneity does not matter at all for culture if no one is engaged in agriculture, and that it matters more the higher the share of farmers. An important caveat for this empirical exercise is that the share of individuals that are engaged in agriculture is potentially a "bad control", especially over a long period, since it might be endogenous to soil heterogeneity. Therefore, estimates from this exercise should not be given a causal interpretation. Instead, they are only meant to provide suggestive evidence regarding likely (or unlikely) channels.

I report the results in Table 3. Column 1 reports my preferred specification for estimating the association between the SHI and the LNI (also in Table 2, column 4). In column 2, I add to the RHS of equation 2 the share of children in-sample that are from farmers' families.[21] Holding the overall share of farming in the county fixed, the estimate of β is slightly lower than in Table 2, but the negative association between soil heterogeneity and close-knit social networks remains significant. Importantly, this result implies that the association does not simply operate through an impact on manufacturing and

[20] See Appendix D.1 for details on the construction of the SHI.

[21] Note that what matters here is not the share of farmers per se, but the share of parents that are farmers.

urbanization (Fiszbein, 2019).

In column 3, I also add the interaction between soil heterogeneity and the share of farmers. The results suggest that, first, the main effect of soil heterogeneity is insignificant, implying that soil heterogeneity is not associated with culture when there is no agricultural activity in the county. This result is reassuring, since if soil heterogeneity only affects culture through farmers' social learning, but no one is engaged in agriculture, soil heterogeneity should not matter for culture. Second, the main effect of the share engaged in agriculture is positive and significant, implying a strong (non-causal) association between the share of farmers in a county and close-knit social networks. This matches the previous findings in the literature, which documented an association between urbanization and individualism (Enke, 2020; Triandis et al., 1990; Vandello and Cohen, 1999). Third, the interaction term is negative and significant, suggesting that soil heterogeneity matters more for culture the more agricultural a county is. This is exactly what we should expect to find if indeed the channel is an impact on farmer's social ties. For the median share of farmers in the sample (40%), an increase from a complete soil homogeneity to a complete soil heterogeneity is associated with about $5.02 \times 0.40 = 2.01$ points decrease in the LNI.

Although not causally identified, those results provide empirical support to Prediction 2 and are highly suggestive of an association between soil heterogeneity and culture that operates through an impact on farmers' attributes and social ties.

3.4.2 Soil Heterogeneity and Agricultural Diversity

According to Shannon's social learning hypothesis, in areas of low soil heterogeneity the optimal farming practices were did not vary across space, which implied that farmers could learn from each other, and their actions quickly converged to the optimum. In areas of high soil heterogeneity, on the other hand, the optimal agricultural practices were highly location specific, and farmers had to individually learn their unique optimum. While an association between local soil heterogeneity and agricultural diversity does not rule out other explanations (see section 2), it is necessary for the social learning hypothesis. In this section I therefore document the relationship between local soil heterogeneity and agriculture diversity.

Using county-level data on the number of acres used in the production of different agricultural products for the years 1880-1935 (Manson et al., 2020), I calculate an agricultural diversity index defined as one minus the the Herfindahl–Hirschman Index.[22] The agricultural diversity index measures the probability that two randomly drawn acres used in farms in a county are used to grow different agricul-

[22] See Appendix D.5 for details on the construction of the index.

tural products. Intuitively, the index is meant to capture the degree to which farmers in the county are cultivating the same agricultural products. An important caveat is that due to the aggregated nature of the data, the diversity captured by the index could be a diversity across acres within a farm rather than diversity across farmers in the county.

Using equation 2, I estimate the association between soil heterogeneity and agricultural diversity. Table 4 reports the results. Across all columns, I find that soil heterogeneity increased agricultural diversity. Depending on the controls includes in the analysis, I find that an increase from SHI = 0 to SHI = 1 increases agricultural diversity by 0.542 to 0.828 standard deviations. This result is reassuring, suggesting that the scope for social learning in a county with a high degree of local soil heterogeneity may have been limited.

3.4.3 Soil Heterogeneity Lowers the Rate of Adoption of New Technology

I also provide suggestive evidence that soil heterogeneity limited the scope of farmers' social learning. I do so by documenting a negative reduced-from association between local soil heterogeneity and the rate of adoption of fertilizers. I use county-level data on the number of farms reporting expenditures on fertilizers for the years 1910-1930 (Manson et al., 2020) to calculate the rate of fertilizers adoption, defined as the growth of the share of farms using fertilizers out of the total number of farms.

Using equation 2, I estimate the association between local soil heterogeneity and the rate of fertilizer adoption. I report the results in Table 5. Across all columns, I find that local soil heterogeneity decreased the growth rate of the share of farms using fertilizers. Depending on the empirical specification, I find that an increase from a complete soil homogeneity to a complete soil heterogeneity decreases the rate of adoption by 0.526 to 0.223 standard deviations. Columns 1-4 are parallel to the columns in Tables 2 and 4, and demonstrate that the result is robust to the inclusion of fixed effects and geo-climatic controls. Column 5 demonstrates that the higher rate of adoption is not explained by a concave adoption function and a lower starting level of adoption. The parameter δ suggested by Oster (2019) is larger from 1 (in absolute value) for all columns 2-5, which suggests that the association between local soil heterogeneity and a lower rate of new technology adoption is unlikely to be explained by an omitted variable bias. The negative association between local soil heterogeneity and the rate of fertilizers adoption is consistent with soil heterogeneity limiting farmers' scope for social learning.

3.4.4 Confounding Channels

Finally, I provide evidence suggesting that the impact of soil heterogeneity on close-knit communities did not operate through two confounding channels: agricultural inequality and birthplace diversity of settlers.

First, it is theoretically possible that a higher degree of local soil heterogeneity will contribute to a higher degree of agricultural inequality due to a higher variation in the agricultural productivity of land. A higher degree of agricultural inequality could, in turn, weaken communal ties and result in a loose-knit social structure.

Second, if settlers were trying to find locations that resembled the agricultural conditions they were familiar with back home, it may be the case that a higher degree of soil heterogeneity will cause a higher degree of birthplace diversity, as different soils may attract settlers from different locations. This in turn might result in a loose-knit social structure. A central consideration working against this story is the fact that soil properties that differentiate soil into different taxonomic classifications can not be determined from the surface alone, and therefore, are not likely to generate a strong selection. Yet this possibility is nevertheless worth exploring.

I test both those possibilities in Table 6 by directly controlling for contemporary birthplace diversity and agricultural inequality, as measured by the Gini coefficient for the distribution of farms' size. As in section 3.4.1 above, those variables are "bad control," as the entire point of this exercise is that they might be endogenous to soil heterogeneity. Therefore, the estimates in columns 2-4 of Table 6 should not be given a causal interpretation. They only serve to provide suggestive evidence on channels.

Column 1 reports the result from my baseline specification of equation 2 (also in Table 2, column 4). Column 2 adds to the estimation equation the Gini coefficient for the distribution of farms' size.[23] The (non-causal) relationship between agricultural inequality and close-knit communities is negative, as expected. However, there is little impact on the estimate of β. The point estimate of β slightly drops (in absolute value) relative to that in column 1, but the difference is insignificant and the estimate of β remains both economically and statistically significant. This suggests that agricultural inequality is not a main driver of the negative association between soil heterogeneity and close-knit communities. Column 3 reports similar patterns with regards to birthplace diversity, suggesting that it too is unlikely to be a main driver of the association between soil heterogeneity and close-knit communities. Finally, column 4 adds to the estimation equation both of the confounding factors. While the change in the point estimate of β relative to the baseline specification is slightly larger in this case, the difference is still insignificant, and here too the estimate of β remains both economically and statistically significant. Those findings suggest that while these two confounding channels may hold some merit, they are far from being a central part of the story.

[23]Data on the distribution of farms' size does not exist for 1850. Therefore the analysis in columns 2 and 4 is carried out over a smaller sample of 1860-1940.

3.5 Alternative Measures

In this section I document that the relationship between local soil heterogeneity and close-knit social structure holds more generally, by using two different proxies for close-knit social structure - religious diversity and the strength of family ties.

3.5.1 Religious Diversity

I use county-level data on the number of members of religious institutions by denomination between 1850-1926 (Manson et al., 2020) to construct a county-level *"Religious Diversity Index"* (RDI), defined as one minus the Herfindahl–Hirschman Index.[24] The index measures the probability that two randomly drawn individuals from the population of members of religious institutions in a county belong to a different denomination. Intuitively, the index captures the degree to which multiple cultural (religious) identities exist within a community in a given year. An important caveat is that the diversity captured by the index could be diversity across different communities within a county rather than diversity within communities in the county.

I study the relationship between soil heterogeneity and the RDI using equation 2. Appendix A presents the spatial variation in the RDI and reports the key results. I find that an increase in soil heterogeneity is associated with an increase of religious diversity. Appendix C.3 documents the robustness of the results to alternative ways to account for spatial auto-correlation and alternative definitions of the SHI.

3.5.2 The Strength of Family Ties

While my main interest is in the effect of soil heterogeneity on the formation of communities with close-knit social networks, it seems plausible that the impact of soil heterogeneity on personality extended more broadly to other forms of social relationships and other in-groups. To explore this possibility, I use data on family structure and the choice of living arrangements from the full count census data between 1860-1940 to construct a county-level measure of the *"Strength of Family Ties"* (SFT).[25,26] Research in social psychology identified family ties as a key factor that correlates with collectivism across cultures (Triandis et al., 1990; Triandis, 2001; Vandello and Cohen, 1999). Similarly, in economics, strong family ties have been shown to correlate with many attributes that are associated with close-knit social networks and interdependence more broadly, such as lower geographical mobil-

[24] See Appendix D.4 for details on the construction of the RDI.

[25] 1850 is excluded because information regarding marital status was not recorded.

[26] See Appendix D.3 for details on the construction of the SFT.

ity, generalized trust, and adverse attitudes toward changes (Alesina and Giuliano, 2010, 2011, 2014; Alesina et al., 2015). SFT has the advantage that, like naming patterns, it is observable in historical census data.

I study the relationship between soil heterogeneity and the SFT using equation 2. Appendix B presents the spatial variation in the SFT and reports the key results. I find that an increase in soil heterogeneity weakens family ties, which is consistent with a negative impact of soil heterogeneity on interdependence more generally. One possibility is that the impact of soil heterogeneity on culture and personality originated with communal ties, but eventually extended to other forms of social relationships and other in-groups. Appendix C.4 documents the robustness of the result to alternative ways to account for spatial auto-correlation and alternative definitions of the SHI.

4 Persistence and Long-Run Impact

After documenting a reduced form relationship between soil heterogeneity and close-knit communities, in this section I study the long-run impact of soil heterogeneity on close-knit social structure. I show that while the historical association between soil heterogeneity and close-knit communities weakens over time, soil heterogeneity is still associated with a lower importance of communal moral values.

4.1 Soil Heterogeneity and Communal Identity, Year-by-Year

The results presented in section 3 studied the association between soil heterogeneity and identification with the local community polling together 1850-1940 in a panel specification. In Table 7, I report results on the association year-by-year instead. The sign on the estimate of β is always negative, and it is statistically significant in almost every period when estimated separately (columns 1-9). However, the strength of the association seems to be weakening over time. The point estimate drops from -4.98 ($p-value < 0.001$) in 1850 (column 1) to -1.00 ($p-value = 0.218$) in 1940 (column 9). Importantly, this pattern is not the result of the change in sample that results from the continuing westward expansion.[27] The decline in magnitude over the years is consistent with soil heterogeneity mattering more in early periods, in which farmers were migrating to new and unfamiliar locations and new communities were forming, and slowly decaying over time. It is also consistent with soil heterogeneity mattering less as the economy was slowly transitioning out of agriculture.

[27] In fact, this pattern is even stronger when I fix the states in sample to the states that existed in 1850.

4.2 The Long-Run

I use contemporary data on individuals' morality to study the long-run impact of soil heterogeneity. The psychologist Haidt (2008) distinguished between two main cultural approaches to morality. The first is an "Individualizing" (or "universal") approach, in which individuals are the fundamental units of moral value, and people are encouraged to respect the rights of others, to stand for universal justice, and to empathize with and care for the weak and vulnerable. The second approach is a "binding" (or "communal") one, in which the group serves as the fundamental source of moral value and individuals are bind together into larger collectives which they are expected to serve. There is a clear analogy between this distinction in the moral domain and the individualism-collectivism cultural cleavage.

Continuing this line of thought, Haidt and his collaborators developed the *"Moral Foundations Theory"* (Haidt and Graham, 2007), which argues that there are five psychological "foundations" of morality, and cultures differ from each other on the degree to which they base their morality on these foundations. Two of the foundations- Harm / Care, and Fairness / Reciprocity, correspond to the individualizing approach to morality, and three- In-group / Loyalty, Authority / Respect, and Purity / Sanctity, to the binding approach. To measure the degree to which individuals' moral judgment involves the five foundations Graham et al. (2011) developed the *"Moral Foundations Questionnaire"* (MFQ) and surveyed on `www.yourmorals.org` approximately 242,000 Americans between 2008-2018.[28,29] To study the association between soil heterogeneity and moral values I match MFQ respondents to contemporary U.S. counties.[30]

4.2.1 The Estimation Framework

I study the relationship between soil heterogeneity and individuals' morality in the following estimation framework:

$$Moral\ Values_{ict} = \beta Soil\ Heterogeneity_c + \theta_{s(c)t} + X_c\Gamma + X_i\Lambda + \epsilon_{ict} \tag{3}$$

where $Moral\ Values_{ict}$ is a measure of the moral values of interest for individual i that resides in county c, which was recorded in year t. $\theta_{s(c)t}$ is a state-by-year fixed effect, X_c is a vector of

[28] For more details on the MFQ data and the five foundations see Appendix D.6.

[29] The use of this data in economics was pioneered by Enke (2020).

[30] I match individuals to county using HUD USPS ZIP code to county crosswalk. Some zip code areas intersect more than one county. In those cases, I match the respondent to all possible counties (i.e. "duplicate" the respondent). In the analysis, I weight each observation by the ratio of residential addresses in the ZIP-county to the total number of residential addresses in the entire ZIP code area (Wilson and Din, 2018), so that each respondent receives a total weight of one.

county level geo-climactic controls, and X_i is a vector of pre-determined individual characteristics fixed effects- age, gender and race. β is the coefficient of interest, representing the relationship between the degree of local soil heterogeneity in i's county of residence and the outcome of interest. To account for spatial auto-correlation I cluster observations at arbitrary grid-cells of size 100 square miles (Bester et al., 2011).

4.2.2 Results

I find that local soil heterogeneity lowers the importance of the binding moral foundations. Table 8 reports the results. The association between soil heterogeneity and the two individualizing factors- Care / Harm, and Fairness / Reciprocity, is indistinguishable from zero (columns 1-2). However, there is a negative association between local soil heterogeneity and all the three binding factors- In-group / Loyalty (column 3), Authority / Respect (column 4) and Purity / Sanctity (column 5). Moreover, the magnitude of the association is comparable across those three factors- an increase from complete soil homogeneity (SHI = 0) to complete soil heterogeneity (SHI = 1) is associated with 0.052-0.069 standard deviation drop in their importance.

Finally, in column 6, I report the estimate when the dependent variable is the first eigenvector from a principal component analysis on the five foundations. I refer to this eigenvector as *"Binding versus Individualizing"*, as it is not just the most important factor, explaining about 46% of the variance in the data, but also the only eigenvector for which the signs of the loadings on the five foundations corresponds to the "binding" versus "individualizing" distinction (positive for binding foundations, negative for individualizing). I find that an increase from a complete soil homogeneity to a complete soil heterogeneity is associated with a 0.064 standard deviation drop in binding versus individualizing moral values.

4.3 Interpretation

This section documented a long-run association between soil heterogeneity and a lower importance of binding moral values. What explains this association?

My preferred interpretation focuses on cultural persistence (Bisin and Verdier, 2001; Guiso et al., 2006; Nunn and Wantchekon, 2011; Voigtländer and Voth, 2012; Alesina et al., 2013). Soil heterogeneity shaped the nature of social relationships of the initial farmers that settled the differed locations across the U.S. Importantly, this impact occurred at a formative period, during which new communities were created, and local institutions and social norms were established. Such an environment constitutes a "critical juncture," making long-lasting effects more plausible.

On the other hand, it seems less likely that those long-run effects can be explained by a continuing impact of soil heterogeneity on the scope for farmers' social learning. Mass migration to settle and farm new locations in the contiguous U.S. had long ended, and contemporary farmers have access to precise and detailed information regarding the nature of soil on their plots. Moreover, only a small fraction of the population is currently engaged in agriculture, making it hard to imagine an impact on the culture of the general population.

An historical impact on culture that persists, rather than a continuing impact, is also consistent with the pattern of a decaying association between soil heterogeneity and close-knit communities what emerges when the data is analysed year-by-year (Table 7). The association was strongest while vast regions of the U.S. were actively settled by farmers, and it got weaker as the frontier was closing and the share of agriculture in the economy declined.

5 The Formation of Close-knit Communities

After documenting a reduced form relationship between soil heterogeneity and close-knit communities and its persistence, I proceed to study the historical formation of this relationship. I present evidence supporting a causal treatment effect interpretation of soil heterogeneity on culture. I focus on the nineteenth century and exploit within-family variation in naming patterns across families that migrated to counties with varying degrees of soil heterogeneity in a Difference-in-Differences framework. I document a decrease in communal identity among farmers that moved to soil-heterogeneous counties relative to farmers that moved to soil-homogeneous counties, along with a null impact on non-farmers moving to similar locations.

To construct the sample, I follow the procedure in Bazzi et al. (2020): Using complete count census data from 1850-1880, I first identify families that migrated within the U.S. using information on the state of birth of their children.[31] I restrict the baseline sample to include families that moved only once, that is, families with children born in exactly two different states, where the later state is also the current state of residence. Then, for each family, I proxy the year of migration using the information on children's age and state of birth. For example, consider a family who lives in Massachusetts 1850, who has two children- the first was born in 1842 in New York and the second in 1846 in Massachusetts. The proxy of the move year for that hypothetical family is 1844. Finally, I assign to each child i a birth year relative to the year in which his/her family moved as $b_i = birth\ year_i - move\ year_{f(i)}$. In the example above, the values are $b_1 = -2$ and $b_2 = 2$, for the first and second child, respectively.

[31] Note that this procedure therefore misses other potential moves across counties within a state.

5.1 Estimation frameworks

Difference-in-Differences. I identify the impact of local soil heterogeneity on parents' name choices in a *Difference-in-Differences* framework, comparing names of children born in families that migrated to counties with varying degrees of local soil heterogeneity, before and after the family had moved.[32] Regression takes the form:

$$LNI_{ibfc} = \delta_b + \theta_{f(i)} + \sum_{b=-5^+}^{7^+} \beta_b \cdot \delta_b \cdot Soil\ Heterogeneity_c + X_i\Omega + \epsilon_{ibfc} \tag{4}$$

where LNI_{ibfc} is the LNI score with "local" defined as the state that is assigned to child i,[33] who was born b years relative to the year his/her family f moved to state $s(c)$, and currently resides in county c. δ_b is a relative-year-of-birth fixed effect, which controls for the dynamics in naming patterns relative to the move year in the baseline of families that moved into completely soil-homogeneous counties (SHI = 0). θ_f is a family fixed effect, which removes all of the variation relating to time-invariant factors at the family level, including the family's cultural background, their economic characteristics at year t, the census decade, and the geo-climatic characteristics of the destination county, including the main effect of soil heterogeneity. X_i is a potential vector of child i characteristics, which I include as a robustness check, and includes gender, birth order, and a 5-year cohort fixed effect. I cluster the standard errors ϵ_{ibfc} at the county-level (Bertrand et al., 2004). The coefficients of interest, β_b, identify the dynamic in naming patterns of children who's family moved to a county with a complete soil heterogeneity (SHI = 1) relative to children who's family moved to completely soil-homogeneous counties. I normalize β_0 to zero so that β_b is interpreted as the effect relative to the year of migration.

The identifying assumption is that absent of soil heterogeneity, children naming patterns of in families that moved into different counties would have evolved similarly after the move. While this assumption can not be verified, I assess its plausibility by studying the pre-trends ($\beta_{b<0}$) in equation 4.

I also estimate the "canonical" Difference-in-Differences specification:

[32] This design resembles the designs in Bazzi et al. (2020) and Abramitzky et al. (2020). The main difference between those designs mine is that I am interested in the differential change in naming habits for families that moved into counties with different levels of soil heterogeneity.

[33] Note that since I only observe the state of birth and not the county of birth, I can not perform the analysis for an LNI in which "local" is defined as the county.

$$LNI_{ibfc} = \delta_b + \theta_{f(i)} + \beta \cdot \mathbb{1}\{b > 0\} \cdot Soil\ Heterogeneity_c + X_i\Omega + \epsilon_{ibfc} \tag{5}$$

My main interest is in estimating equations 4 and 5 *for farmers' families*, as this is the group for which, according to the social learning hypothesis and Prediction 2, should be directly affected by soil heterogeneity. However, I report heterogeneous treatment results- the estimation for all families, farmers' families, and non-farmers' families, since the comparison of effects across those groups is compelling.

Triple-Difference. Motivated by the heterogeneous treatment analysis across farmers' and non-farmers' families, I also estimate the following *Triple-Difference* specification:

$$\begin{aligned} LNI_{ibfc} =& \delta_b + \theta_{f(i)} + \delta_b \cdot \mathbb{1}\{farmer_f\} + \\ & \sum_{b=-5^+}^{7^+} \gamma_b \cdot \delta_b \cdot Soil\ Heterogeneity_c + \\ & \sum_{b=-5^+}^{7^+} \beta_b \cdot \delta_b \cdot \mathbb{1}\{farmer_f\} \cdot Soil\ Heterogeneity_c + X_i\Omega + \epsilon_{ibfc} \end{aligned} \tag{6}$$

In this specification, the coefficients of interest, β_b, identify the dynamic in naming patterns of farmers' children relative to non-farmers' children in families that moved to a county with complete soil heterogeneity (SHI = 1) relative to the same difference for families that moved to completely soil-homogeneous counties (SHI = 0). β_0 is normalized to equal zero.

Note that the identification assumption in this specification is different than in the Difference-in-Differences specification. It requires that in the absence of soil heterogeneity, the trend in children naming patterns of farmers' and non-farmers' families would have evolved similarly after the move across different counties.

Finally, I also estimate a "canonical" Triple-Difference specification:

$$\begin{aligned} LNI_{ibfc} =& \delta_b + \theta_{f(i)} + \mathbb{1}\{b > 0\} \cdot \mathbb{1}\{farmer_f\} + \\ & \gamma \cdot \mathbb{1}\{b > 0\} \cdot Soil\ Heterogeneity_c + \\ & \beta \cdot \mathbb{1}\{b > 0\} \cdot \mathbb{1}\{farmer_f\} \cdot Soil\ Heterogeneity_c + X_i\Omega + \epsilon_{ibfc} \end{aligned} \tag{7}$$

5.2 Results

Figure 3 presents the estimates for equations 4 and 6. In Figure 3(A), I present the estimates of β_b from equation 4 when the sample includes all families. The lack of trend in LNI prior to migration suggests that there was no selective migration on prior levels of communal identification to counties with varying degrees of soil heterogeneity. After the move, there is a sharp decline in LNI in high soil heterogeneity counties relative soil-homogeneous counties. However, the precision of the estimates is low. The effect peaks (in absolute value) three years after the move, before partially reverting toward the pre-move levels of differences. The LNI of a child born to a family three years after they had moved to a county with a complete soil heterogeneity (SHI = 1) was about 2.82 ($p-value = 0.058$) points lower relative to the LNI of a child born to a family who moved to a county with a complete soil homogeneity (SHI = 1).

Figure 3(B) plots the estimates of β_b from the same equation when the sample only includes farmers' families. The dynamic in LNI is similar to the dynamic estimated using the full sample, however, both magnitudes and precision are higher. The LNI of a child born to a farmers' family one years after a move to a county with SHI = 1 was about 2.92 ($p-value = 0.079$) points lower relative to that of a child born to farmers who moved to a county with SHI = 0. The impact intensified in the next couple of year, with a 3.46 ($p-value = 0.047$) points decrease in the second year and a 4.57 ($p-value = 0.009$) points decrease in the third year. In the fourth year after the move to a high local soil heterogeneity county the effect attenuates to a 3.41 ($p-value = 0.066$) points decrease before flattening out at about 2.5 point decrease starting from the fifth year, although insignificant.

The quick response of children's LNI to soil heterogeneity also provides support the social learning hypothesis. For newly migrated farmers the first few years were the most important period for learning. During this period farmers located in an environment in which social learning was possible had a strong incentive to socialize with their neighbors and pay close attention to their actions. After the first few years, when framers had learned the optimal framing practices on their new land, those incentives should have diminished. The dynamic reveled in Figure 3(B) - a discontinuous "jump" in the first year, with a subsequent gradual strengthening of the effect over the a few years, followed by a partial reversion, therefore matches the social learning interpretation. At the same time, this pattern does not seem to match the expected dynamic of some of the alternative channels. For example, if the impact is due to similarly of action that over time contributed to the development of a "tight" culture and a strong emphasis on norms, then we would expect the effect to slowly pickup over time.

Figure 3(C) plots the DiD estimates when the sample only includes non-farmers' families and reveals no differential impact for families that moved to counties with different levels of soil heterogeneity.

While the precision of the DiD estimates are generally not very high, contrasting the estimates for farmers with those of non-farmers is compelling. A causal impact of soil heterogeneity on farmers' attributes along with a null impact on the attributes of non-farmer, in line with Prediction 2, provides strong evidence in favor of the social learning hypothesis. Those finding substantially reduce the likelihood that the reduced-form relationship between local soil heterogeneity and close-knit communities documented in section 3.3 is explained by factors entirely different than farmers' direct exposure to a high versus low soil heterogeneity environment.

To drive the point home, I also report the estimates of the Triple-Difference specification (equation 6) in Figure 3(D). Before moving, there is no dynamic in the difference between the LNI of farmers' children relative to non-farmers' children in high soil heterogeneity counties and the LNI of farmers' children relative to non-farmers' children in low soil heterogeneity counties. However, after the move, the differences between the LNI of farmers' and non-farmers' children grows in counties with high soil heterogeneity relative to low soil heterogeneity counties. Three years after the move, the triple-difference estimate reaches $\beta_3 = -5.09$ ($p-value = 0.088$) before partially reverting. While statistical power in this exercise is relatively low, the trend documenting differential response to soil heterogeneity within a county is striking.

Table 9 reports the estimates from the baseline "canonical" Difference-in-Differences design (equation 5, columns 1-3) and the Triple-Difference design (equation 7, column 4). The findings are consistent with the estimates from the dynamic specifications, however, the aggregation of the *Post move* effects increases precision. The sample in column 1 includes all families. It suggests that the LNI of children born to families following a move to a county with a complete soil heterogeneity was 1.83 ($p-value = 0.029$) points lower relative to children who were born following a move to a fully soil-homogeneous county. In columns 2 and 3, I estimate equation 5 separately for farmers' and non-farmers' families, respectively. Column 2 documents a strong impact of soil heterogeneity on the culture of farmers. An increase in soil heterogeneity from SHI = 0 to SHI = 1 decreases the LNI score of children born after the move by 3.25 ($p-value < 0.001$) points. Column 3 documents the null impact on non-farmers' families. Finally, in column 4, I report the estimate of the canonical Triple-Difference design, which suggests that the difference between the LNI score of farmers' and non-farmers' children born after migration to a county with complete soil heterogeneity was 4.02 ($p-value < 0.001$) points lower relative to the same difference among families that moved to a soil-homogeneous county.

Robustness. The Difference-in-Differences results in Table 9 are robust to adding individual controls and to alternative sample construction. Appendix Table C.3 and Appendix Figure C.2 document the robustness of the results to controlling for child's gender, birth order and a 5-year cohort fixed effect.

Appendix Table C.4 documents robustness to also including in the sample families with foreign-born parents (Panel A), non-white families (Panel B), or families that moved more than once (Panel C).

6 Concluding Remarks

This paper studies the cultural implications of social learning. During the settlement of the United States, millions of farmer migrated to new and unfamiliar environments. To survive, they had to quickly learn the optimal location-specific farming practices. In some locations, soil was relatively homogeneous and farmers could learn from their neighbors, while in other locations high degree of soil heterogeneity limited the scope for social learning. Those differences materialized into differences in the cultural importance attached to the community. I show that historically, soil-heterogeneous counties displayed weaker communal ties. I go beyond correlation by providing causal evidence on the formation of this pattern. I document a decrease in the strength of communal ties of farmers following migration to a soil-heterogeneous county. In contrast, the social ties of non-farmers were not affected by soil heterogeneity. This pattern provides further support for a social learning interpretation of the reduced-form association between soil heterogeneity and culture. Today, soil-heterogeneous counties continue to be less communal.

The distinction between close-knit and loose-knit cultures is often considered to be the fundamental cultural cleavage across human societies. This cleavage is the focus of considerable multidisciplinary research across the social sciences. Studies in economics have related it to economics growth (Gorodnichenko and Roland, 2011, 2017) and contemporary political preferences (Graham et al., 2009; Enke, 2020; Enke et al., 2019). Understanding the deep roots of this cleavage is therefore meaningful and important. I provide the first direct empirical evidence supporting the *"Social Learning Hypothesis"*, put forth by the historian Fred Shannon 75 years ago, but received little to no attention since. The findings of this paper suggest that, while understudied, social learning is an important factor in the formation of close-knit communities.

This study opens up an avenue for future research on the cultural implications of social learning. The empirical evidence provided in this paper is historical and reduced form in nature. Other research designs, such as experimental designs, may shad light on the specific mechanisms in play. Specifically, it may help to distinguish between the different channels by which social learning can affects the strength of social ties, such as an impact on the incentives to invest in social relationships versus a direct impact on personal believes and attitudes.

References

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson, "Do Immigrants Assimilate More Slowly Today Than in the Past?," *American Economic Review: Insights*, March 2020, *2* (1), 125–41.

Alesina, Alberto and Paola Giuliano, "The power of the family," *Journal of Economic growth*, 2010, *15* (2), 93–125.

_ and _ , "Family ties and political participation," *Journal of the European Economic Association*, 2011, *9* (5), 817–839.

_ and _ , "Family ties," in "Handbook of economic growth," Vol. 2, Elsevier, 2014, pp. 177–215.

_ , _ , and Nathan Nunn, "On the origins of gender roles: Women and the plough," *The Quarterly Journal of Economics*, 2013, *128* (2), 469–530.

_ , Yann Algan, Pierre Cahuc, and Paola Giuliano, "Family values and the regulation of labor," *Journal of the European Economic Association*, 2015, *13* (4), 599–630.

Ang, James B, "Agricultural legacy and individualistic culture," *Journal of Economic Growth*, 2019, *24* (4), 397–425.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse, "Frontier culture: The roots and persistence of "Rugged Individualism" in the United States," *Econometrica, Forthcoming*, 2020.

Beck-Knudsen, Anne Sofie, "Those who stayed: Selection and cultural change during the age of mass migration," *Unpublished Manuscript*, 2019.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-In-Differences Estimates?," *The Quarterly Journal of Economics*, 02 2004, *119* (1), 249–275.

Besley, Timothy and Anne Case, "Diffusion as a Learning Process: Evidence from HYV Cotton," *Princeton, Woodrow Wilson School - Development Studies*, 1994, (174).

Bester, Alan C., Timothy G. Conley, and Christian B. Hansen, "Inference with Dependent Data using Cluster Covariance Estimators," *Journal of Econometrics*, 2011, *165* (2), 137–151.

Bisin, Alberto and Thierry Verdier, "The economics of cultural transmission and the dynamics of preferences," *Journal of Economic theory*, 2001, *97* (2), 298–319.

Boyd, Robert and Peter J Richerson, *Culture and the evolutionary process*, University of Chicago press, 1988.

Buggle, Johannes C, "Growing collectivism: Irrigation, group conformity and technological divergence," *Unpublished Manuscript*, 2018.

Chang, Lei, Miranda CK Mak, Tong Li, Bao Pei Wu, Bin Bin Chen, and Hui Jing Lu, "Cultural adaptations to environmental variability: An evolutionary account of East–West differences," *Educational Psychology Review*, 2011, *23* (1), 99–129.

Conley, Timothy, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92*, 1–45.

Conley, Timothy G and Christopher R Udry, "Learning about a new technology: Pineapple in Ghana," *American economic review*, 2010, *100* (1), 35–69.

Ellison, Glenn and Drew Fudenberg, "Rules of thumb for social learning," *Journal of political Economy*, 1993, *101* (4), 612–643.

Enke, Benjamin, "Kinship, cooperation, and the evolution of moral systems," *The Quarterly Journal of Economics*, 2019, *134* (2), 953–1019.

_ , "Moral values and voting," *Journal of Political Economy*, 2020, *128* (10), 3679–3729.

_ **, Ricardo Rodríguez-Padilla, and Florian Zimmermann**, "Moral Universalism and the Structure of Ideology," *Unpublished Manuscript*, 2019.

Fincher, Corey L, Randy Thornhill, Damian R Murray, and Mark Schaller, "Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism," *Proceedings of the Royal Society B: Biological Sciences*, 2008, *275* (1640), 1279–1285.

Fiszbein, Martin, "Agricultural Diversity, Structural Change and Long-Run Development: Evidence from the U.S.," *Unpublished Manuscript*, 2019.

Foster, Andrew D and Mark R Rosenzweig, "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of political Economy*, 1995, *103* (6), 1176–1209.

Fouka, Vasiliki, "Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I," *The Review of Economic Studies*, 05 2019, *87* (1), 204–239.

Fryer, Roland and S. Levitt, "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics*, 2004, *119* (3), 767–805.

Galor, Oded and Ömer Özak, "The agricultural origins of time preference," *American Economic Review*, 2016, *106* (10), 3064–3103.

Giuliano, Paola and Nathan Nunn, "Understanding Cultural Persistence and Change," *Review of Economic Studies, Forthcoming*, 2020.

Gorodnichenko, Yuriy and Gerard Roland, "Which dimensions of culture matter for long-run growth?," *American Economic Review*, 2011, *101* (3), 492–98.

_ **and** _ , "Culture, institutions and democratization," *NBER Working Paper*, 2015.

_ **and** _ , "Culture, institutions, and the wealth of nations," *Review of Economics and Statistics*, 2017, *99* (3), 402–416.

Graham, Jesse, Brian A Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H Ditto, "Mapping the moral domain.," *Journal of personality and social psychology*, 2011, *101* (2), 366.

_ **, Jonathan Haidt, and Brian A Nosek**, "Liberals and conservatives rely on different sets of moral foundations.," *Journal of personality and social psychology*, 2009, *96* (5), 1029.

Greenfield, Patricia M, "Linking social change and developmental change: shifting pathways of human development.," *Developmental psychology*, 2009, *45* (2), 401.

Griliches, Zvi, "Hybrid corn: An exploration in the economics of technological change," *Econometrica*, 1957, pp. 501–522.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Does culture affect economic outcomes?," *Journal of Economic perspectives*, 2006, *20* (2), 23–48.

Haidt, Jonathan, "Morality," *Perspectives on Psychological Science*, 2008, *3* (1), 65–72.

_ **and Jesse Graham**, "When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize," *Social Justice Research*, 2007, *20* (1), 98–116.

Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov, *Cultures and Organizations: Software of the Mind. Intercultural Cooperation and Its Importance for Survival*, revised and expanded 3rd ed., New York: McGraw Hill, 2010.

IIASA/FAO, "Global Agro-ecological Zones (GAEZ v3.0)," 2012. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Kitayama, Shinobu, Keiko Ishii, Toshie Imada, Kosuke Takemura, and Jenny Ramaswamy, "Voluntary settlement and the spirit of independence: Evidence from Japan's" northern frontier."," *Journal of personality and social psychology*, 2006, *91* (3), 369.

Lehner, Bernhard, Kris Verdin, and Andy Jarvis, "New global hydrography derived from spaceborne elevation data," *EOS, Transactions, American Geophysical Union*, 2008, *89* (10), 93–94.

Leip, Dave, "David Leip's atlas of U.S. Presidential elections, datasets," 2017. Harvard Dataverse.

Markus, Hazel R and Shinobu Kitayama, "Culture and the self: Implications for cognition, emotion, and motivation.," *Psychological review*, 1991, *98* (2), 224.

Mattheis, Ross and Itzchak T. Raz, "There's No Such Thing As Free Land: The Homestead Act and Economic Development," *Unpublished Manuscript*, 2019.

Mesoudi, Alex, Lei Chang, Keelin Murray, and Hui Jing Lu, "Higher frequency of social learning in China than in the West shows cultural variation in the dynamics of cultural evolution," *Proceedings of the Royal Society B: Biological Sciences*, 2015, *282* (1798), 20142209.

Minnesota Population Center, "Integrated Public Use Microdata Series, International: Version 7.2 [dataset]," 2019. Minneapolis, MN: IPUMS.

Munshi, Kaivan, "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of development Economics*, 2004, *73* (1), 185–213.

Nunn, Nathan and Leonard Wantchekon, "The slave trade and the origins of mistrust in Africa," *American Economic Review*, 2011, *101* (7), 3221–52.

Oishi, Shigehiro, Janetta Lun, and Gary D Sherman, "Residential mobility, self-concept, and positive affect in social interactions.," *Journal of personality and social psychology*, 2007, *93* (1), 131.

_ , Keiko Ishii, and Janetta Lun, "Residential mobility and conditionality of group identification," *Journal of Experimental Social Psychology*, 2009, *45* (4), 913–919.

Olivetti, Claudia and M. Daniele Paserman, "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940," *American Economic Review*, August 2015, *105* (8), 2695–2724.

Olmstead, Alan L and Paul W Rhode, *Creating Abundance*, Cambridge University Press, 2008.

_ and _ , "Adapting North American wheat production to climatic challenges, 1839–2009," *Proceedings of the National Academy of sciences*, 2011, *108* (2), 480–485.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Raz, Itzchak T., "Use It Or Lose It: Adverse Possession and Economic Development," *Unpublished Manuscript*, 2018.

Russo, Gianluca, "Mass Media and Cultural Homogenization: Broadcasting the American Dream on the Radio," *Unpublished Manuscript*, 2019.

Schwartz, Shalom, *Beyond Individualism/Collectivism: New Cultural Dimensions of Values*, Vol. 18, 01

Shannon, Fred A., *The Farmer's Last Frontier: Agriculture, 1860-1897*, New York: J. J. Little and Ives Company, 1945.

Smith, Cory, "Land Concentration and Long-Run Development: Evidence from the Frontier United States," *Unpublished Manuscript*, 2019.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, *Soil Taxonomy. A basic system of soil classification for making and interpreting soil surveys* Agriculture Handbook, Number 436, Second ed. 1999.

_ , "Introduction to Soils, Soils 101," *Web Soil Survey* 2017. Available online https://websoilsurvey.sc.egov.usda.gov/. Accessed [05/25/2017].

_ , *U.S. General Soil Map (STATSGO2)* 2017. Available online https://data.nal.usda.gov/dataset/united-states-general-soil-map-statsgo2. Accessed [06/09/2017].

Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles, "IPUMS National Historical Geographic Information System: Version 15.0 [dataset]," 2020. Minneapolis, MN: IPUMS.

Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek, "IPUMS USA: Version 10.0 [dataset]," 2020. Minneapolis, MN: IPUMS.

Talhelm, T, X Zhang, S Oishi, C Shimin, D Duan, X Lan, S Kitayama et al., "Large-scale psychological differences within China explained by rice versus wheat Agriculture.," *Science (Washington)*, 2014, *344* (6184), 603–608.

Triandis, Harry C, *New directions in social psychology. Individualism & collectivism*, Boulder, CO: Westview Press, 1995.

_ , "Individualism-collectivism and personality," *Journal of personality*, 2001, *69* (6), 907–924.

_ , Christopher McCusker, and C Harry Hui, "Multimethod probes of individualism and collectivism.," *Journal of personality and social psychology*, 1990, *59* (5), 1006.

Turner, Frederick Jackson, "The Significance of the Frontier in American History," in "The Frontier in American History," New York: Henry Holt and Company, 1921.

Vandello, Joseph A and Dov Cohen, "Patterns of individualism and collectivism across the United States.," *Journal of personality and social psychology*, 1999, *77* (2), 279.

Varnum, Michael EW and Shinobu Kitayama, "What's in a name? Popular names are less common on frontiers," *Psychological science*, 2011, *22* (2), 176–183.

Voigtländer, Nico and Hans-Joachim Voth, "Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany," *The Quarterly Journal of Economics*, 2012, *127* (3), 1339–1392.

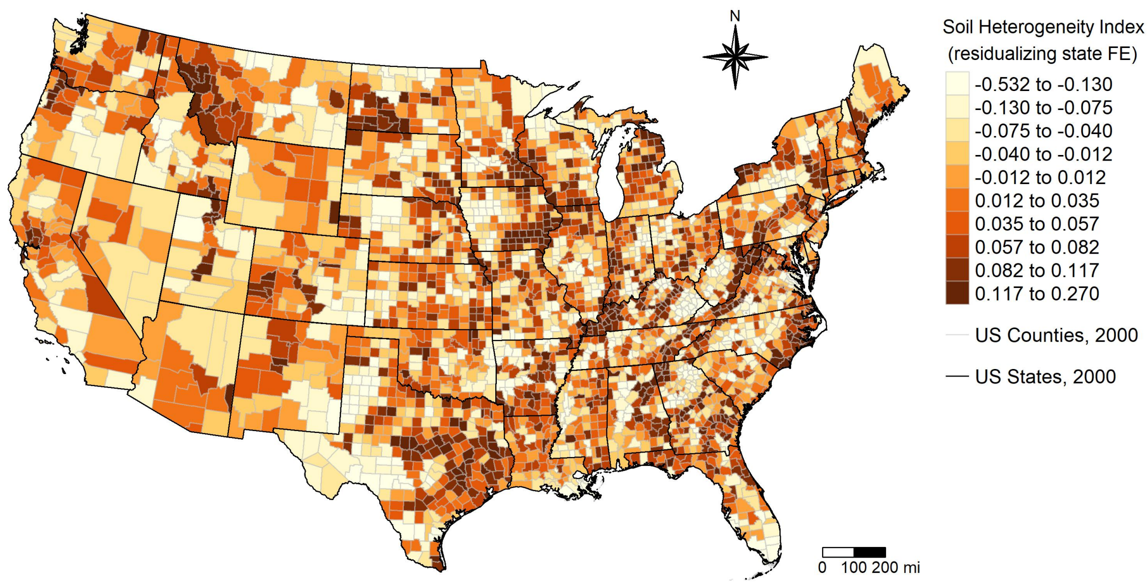
Wilson, Ron and Alexander Din, "Understanding and Enhancing the U.S. Department of Housing and Urban Development's ZIP Code Crosswalk Files," *Cityscape: A Journal of Policy Development and Research*, 2018, *20* (2), 277–294.

Yamauchi, Futoshi, "Social learning, neighborhood effects, and investment in human capital: Evidence from Green-Revolution India," *Journal of development Economics*, 2007, *83* (1), 37–62.

Yaveroglu, Idil Sayrac and Naveen Donthu, "Cultural influences on the diffusion of new products," *Journal of International Consumer Marketing*, 2002, *14* (4), 49–63.

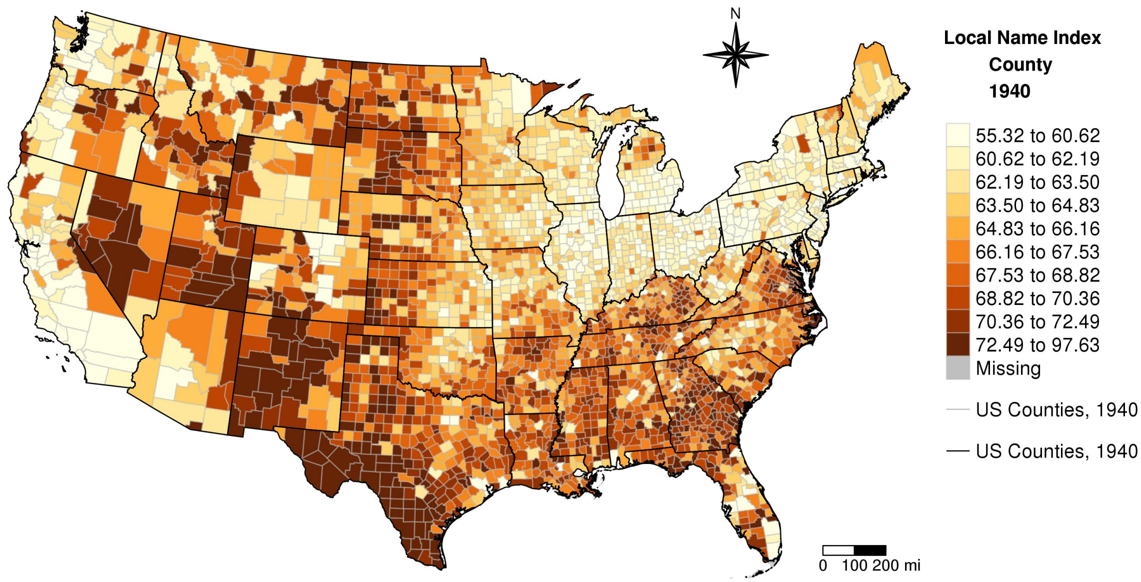
Figures

FIGURE 1: COUNTY-LEVEL SOIL HETEROGENEITY INDEX, 2000



Note: This figure plots the soil heterogeneity index (SHI) for counties in the contiguous U.S. in 2000. State fixed effects are partialed out. Darker color implies a higher soil heterogeneity. See Appendix D.1 for a description of the SHI construction.

FIGURE 2: COUNTY-LEVEL LOCAL NAME INDEX, 1940



Note: This figure plots the county-level *"Local Name Index"* (LNI) in which "local" is defined as the county, for counties in the contiguous U.S. in 1940. Data includes native-born white children between the age of 0 to 10 with native-born parents in the 1940 Census. Darker color implies a higher local name index. See Appendix D.2 for a description of the LNI construction.

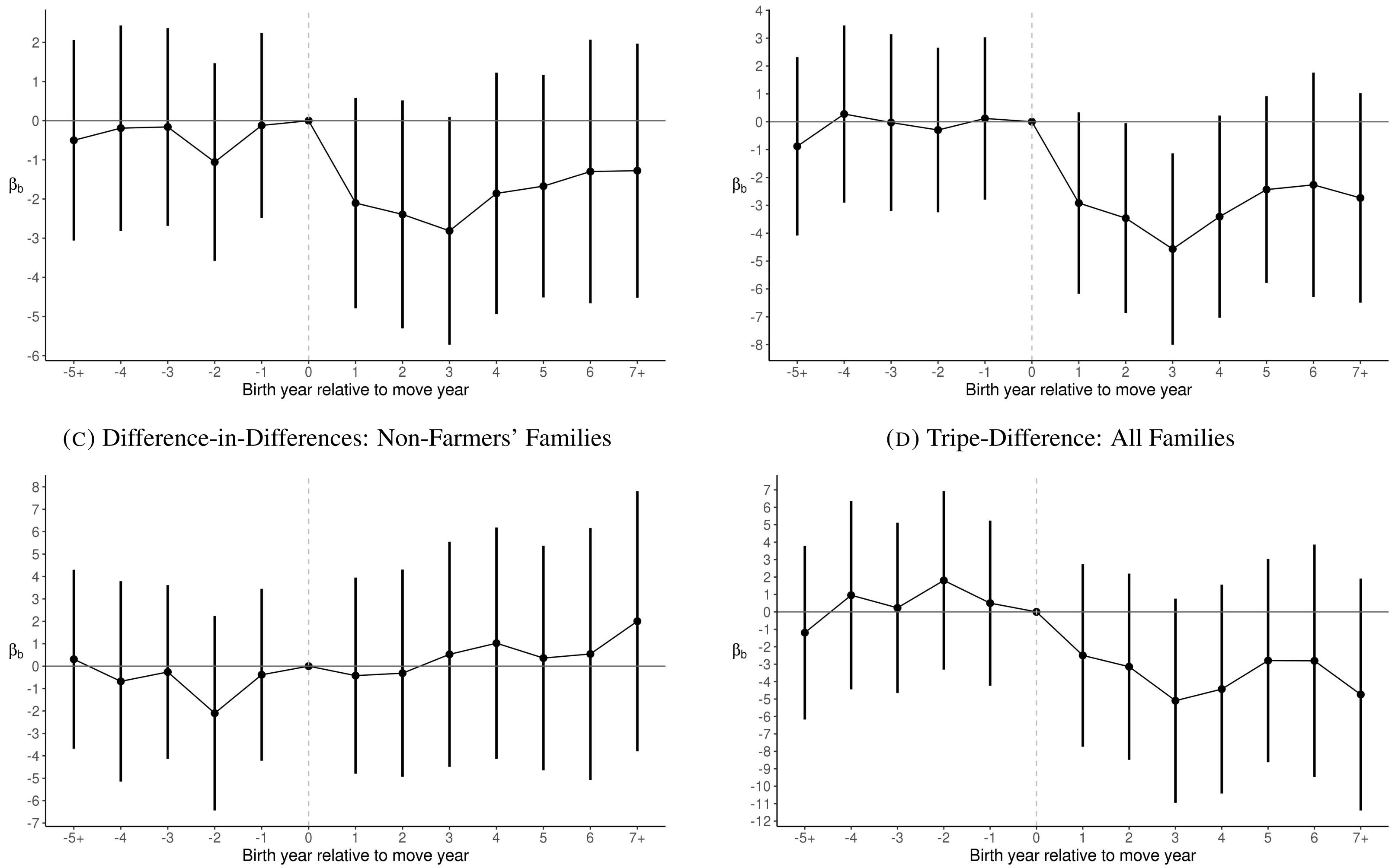
FIGURE 3: IDENTIFYING FARMERS' CULTURAL RESPONSE TO SOIL HETEROGENEITY

(A) Difference-in-Differences: All Families

(B) Difference-in-Differences: Farmers' Families

(C) Difference-in-Differences: Non-Farmers' Families

(D) Tripe-Difference: All Families



Note: This figure plots the estimates of β_b and 95% confidence intervals from equations 4 and 6 when the dependent variable is children's LNI where "local" is defined as the state. The data is from the full count censuses between 1850-1880 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at the contemporaneous county.

Tables

TABLE 1: LOCAL NAME INDEX CORRELATES WITH COMMUNAL MORALITY

	Dependent variable:					
	Rel. Importance of Communal Moral Values		Trump Vote Share 2016		$\Delta\,[Trump - GOP]$	
	(1)	(2)	(3)	(4)	(5)	(6)
Local Name Index	0.030^{***}	0.014^{***}	0.050^{***}	0.044^{***}	0.009	0.050^{***}
	(0.005)	(0.005)	(0.009)	(0.010)	(0.007)	(0.007)
Number of Observations	2,236	2,236	3,085	3,085	3,085	3,085
Number of Clusters	312	312	337	337	337	337
R^2	0.023	0.1	0.063	0.35	0.0022	0.43
State Fixed Effects		✓		✓		✓

Note: This table reports estimates of from regressions in which the independent variable is the 1940 children's LNI in which "local" is defined as the county and the dependent variables are indicators of communal morality from Enke (2020), standardized into z-scores. The data used to calculate the LNI is from the 1940 full count census (Minnesota Population Center, 2019). Data on county-level relative importance of communal values is from Enke (2020). Data on county-level presidential election vote share is from Leip (2017). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

TABLE 2: SOIL HETEROGENEITY REDUCES COMMUNAL IDENTIFICATION

	Dependent variable: Local Name Index			
	(1)	(2)	(3)	(4)
Soil Heterogeneity	-4.524^{***}	-5.511^{***}	-2.914^{***}	-2.486^{***}
	(1.342)	(0.893)	(0.731)	(0.725)
Oster δ for $\beta = 0$		-153.79	2.82	2.52
Number of Observations	23,435	23,435	23,435	23,435
Number of Counties	3,499	3,499	3,499	3,499
Number of Clusters	338	338	338	338
R^2	0.0063	0.47	0.53	0.55
Dependent Variable Mean	67.82	67.82	67.82	67.82
Dependent Variable SD	6.30	6.30	6.30	6.30
State-by-Year Fixed Effects		✓	✓	✓
Geoclimatic Controls			✓	✓
Smooth Location Controls				✓

Note: This table reports estimates of equation 2 when the dependent variable is children's LNI in which "local" is defined as the county. The data is from the full count censuses between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 3: SOIL HETEROGENEITY AFFECTS FARMERS

	Dependent variable: Local Name Index		
	(1)	(2)	(3)
Soil Heterogeneity	−2.486***	−1.426**	0.719
	(0.725)	(0.577)	(1.155)
Share Farmers		9.143***	12.235***
		(0.497)	(1.567)
Soil Heterogeneity × Share Farmers			−5.021**
			(2.381)
Number of Observations	23,435	23,412	23,412
Number of Counties	3,499	3,498	3,498
Number of Clusters	338	338	338
R^2	0.55	0.59	0.59
Dependent Variable Mean	67.82	67.83	67.83
Dependent Variable SD	6.30	6.31	6.31
State-by-Year Fixed Effects	✓	✓	✓
Geoclimatic Controls	✓	✓	✓
Smooth Location Controls	✓	✓	✓

Note: This table reports estimates of equation 2 with an additional control for the share of farmers (column 2) and additionally for the interaction between the share of farmers and soil heterogeneity (column 3). The dependent variable is children's LNI in which "local" is defined as the county. The data is from the full count censuses between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 4: SOIL HETEROGENEITY INCREASES AGRICULTURE DIVERSITY

	Dependent variable: Agriculture Diversity			
	(1)	(2)	(3)	(4)
Soil Heterogeneity	0.735**	0.828***	0.568**	0.542**
	(0.357)	(0.259)	(0.241)	(0.253)
Oster δ for $\beta = 0$		35.31	4.89	4.45
Number of Observations	23,254	23,254	23,254	23,254
Number of Counties	3,337	3,337	3,337	3,337
Number of Clusters	338	338	338	338
R^2	0.0066	0.34	0.37	0.37
State-by-Year Fixed Effects		✓	✓	✓
Geoclimatic Controls			✓	✓
Smooth Location Controls				✓

Note: This table reports estimates of equation 2 when the dependent variable is a county-level agricultural diversity index for the years 1880-1935, standardized into z-scores. The data is from (Manson et al., 2020). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 5: SOIL HETEROGENEITY DECREASES THE RATE OF FERTILIZERS ADOPTION

	Dependent variable: Growth in Fertilizers Use				
	(1)	(2)	(3)	(4)	(5)
Soil Heterogeneity	-0.371^{**}	-0.526^{***}	-0.287^{***}	-0.280^{***}	-0.223^{**}
	(0.152)	(0.126)	(0.103)	(0.101)	(0.095)
Share Using Fertilizer$_{t-1}$					-0.443^{***}
					(0.062)
Oster δ for $\beta = 0$		-11.49	4.33	4.41	2.89
Number of Observations	8,751	8,751	8,751	8,751	8,751
Number of Counties	3,036	3,036	3,036	3,036	3,036
Number of Clusters	336	336	336	336	336
R^2	0.0017	0.19	0.21	0.22	0.22
State-by-Year Fixed Effects		✓	✓	✓	✓
Geoclimatic Controls			✓	✓	✓
Smooth Location Controls				✓	✓

Note: This table reports estimates of equation 2 when the dependent variable is the county-level growth rate of share of farms using fertilizer, standardized into z-scores. The data is from (Manson et al., 2020). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 6: CONFOUNDING CHANNELS

	Dependent variable: Local Name Index			
	(1)	(2)	(3)	(4)
Soil Heterogeneity	-2.486^{***}	-2.072^{***}	-2.211^{***}	-1.827^{**}
	(0.725)	(0.750)	(0.721)	(0.749)
Farms' size Gini		-0.821^{***}		-0.758^{***}
		(0.106)		(0.109)
Birth Place Diversity			-1.139^{***}	-1.174^{***}
			(0.143)	(0.149)
Number of Observations	23,435	21,602	23,435	21,602
Number of Counties	3,499	3,417	3,499	3,417
Number of Clusters	338	338	338	338
R^2	0.55	0.54	0.56	0.54
Dependent Variable Mean	68	68	68	68
Dependent Variable SD	6.3	5.9	6.3	5.9
State-by-Year Fixed Effects	✓	✓	✓	✓
Geoclimatic Controls	✓	✓	✓	✓
Smooth Location Controls	✓	✓	✓	✓

Note: This table reports estimates of equation 2 with additional controls for the Gini coefficient on the distribution of farm sizes (columns 2 and 4) and birthplace diversity index (columns 3-4). The dependent variable is children's LNI in which "local" is defined as the county. The data is from the full count censuses between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Soil Heterogeneity and Communal Identification, Period-by-Period

	Dependent variable: Local Name Index								
Year	1850	1860	1870	1880	1900	1910	1920	1930	1940
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Soil Heterogeneity	−4.978***	−5.717***	−4.188***	−3.801***	−2.078**	−1.814**	−1.783**	−0.621	−1.002
	(1.199)	(1.584)	(1.215)	(1.086)	(1.003)	(0.892)	(0.870)	(0.800)	(0.813)
Number of Observations	1,606	2,031	2,242	2,526	2,819	2,949	3,065	3,098	3,099
Number of Clusters	176	213	261	297	327	328	334	337	337
R^2	0.65	0.56	0.62	0.61	0.57	0.5	0.52	0.47	0.5
Dependent Variable Mean	65.94	69.68	69.44	68.82	68.46	67.84	67.32	66.78	66.54
Dependent Variable SD	6.49	8.16	7.84	7.56	5.97	5.14	5.22	4.93	5.02
State Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geoclimatic Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Smooth Location Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: This table reports estimates of equation 2 when the dependent variable is children's LNI in which "local" is defined as the county, estimated separately for each census year. The data is from the full count censuses between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 8: SOIL HETEROGENEITY WEAKENS LONG-RUN BINDING MORAL VALUES

	Dependent variable:					
	Individualizing		Binding			Binding versus Individualizing
	Care / Harm	Fairness / Reciprocity	In-group / Loyalty	Authority / Respect	Purity / Sanctity	
	(1)	(2)	(3)	(4)	(5)	(6)
Soil Heterogeneity	−0.028	0.006	−0.052*	−0.059**	−0.069**	−0.064**
	(0.017)	(0.017)	(0.027)	(0.026)	(0.029)	(0.032)
Number of Observations	293,663	293,157	293,792	294,015	293,400	272,695
Number of Counties	1,762	1,762	1,762	1,762	1,762	1,762
Number of Clusters	622	622	622	622	622	622
R^2	0.11	0.04	0.03	0.038	0.056	0.043
State-by-Year Fixed Effects	✓	✓	✓	✓	✓	✓
Geoclimatic Controls	✓	✓	✓	✓	✓	✓
Smooth Location Controls	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓

Note: This table reports estimate of equation 3 when the dependent variables are measures of the importance of different moral foundations in individuals' morality, standardized into z-scores. Individual controls include age, gender and race fixed-effects. The data is from the individual responses to the MFQ (Graham et al., 2011) surveyed on `www.yourmorals.org` between 2008-2018. See Appendix D for full details on each dependent variable. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 9: IDENTIFYING FARMERS' CULTURAL RESPONSE TO SOIL HETEROGENEITY

	Dependent variable: Local Name Index			
	Difference-in-Differences			Triple-Difference
Sample:	All Households	Farmer′s Households	Non-Farmer′s Households	All Households
	(1)	(2)	(3)	(4)
Post Move × Soil Heterogeneity	−1.826**	−3.245***	0.772	0.773
	(0.836)	(0.889)	(1.165)	(1.149)
Post Move × Farmers′ Household × Soil Heterogeneity				−4.017***
				(1.179)
Number of Observations	1,203,908	713,881	490,022	1,203,903
Number of Counties	2,559	2,472	2,477	2,559
R^2	0.37	0.35	0.39	0.37
Dependent Variable Mean	54.22	54.40	53.95	54.22
Dependent Variable SD	13.66	13.42	14.00	13.66
Households Fixed Effects	✓	✓	✓	✓

Note: This table presents estimates of equations 5 and 7 when the dependent variable is children's LNI where "local" is defined as the state. The data is from the full count censuses between 1850-1880. See Appendix D.2 for a description of the LNI construction. Standard errors clustered by county in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix

Content

A Religious Diversity

This section studies the relationship between local soil heterogeneity *"Religious Diversity Index"* (RDI). The index measures the probability that two randomly drawn individuals from the population of members of religious institutions in a county belong to a different denomination. Intuitively, the index captures the degree to which multiple cultural (religious) identities exist within a community. I focus on RDI as a second alternative measure for close-knit social networks due to the centrality of religion in personal and communal identity. Intuitively, the existence of many religious identities within the community implies a loose-knit social structure at the community level.

To construct the index, I use county-level data on the number of members of religious institutions by denomination between 1850-1926 (Manson et al., 2020). The RDI is defined as one minus the Herfindahl–Hirschman Index over the share of members of religious denominations. Formally:

$$Religious\ Diversity\ Index_{ct} = 1 - \sum_j s_{cjt}^2$$

where s_{cjt} is the share of members of religious denomination j in county c in year t out of the total number of members in religious institutions in county c year t. For more details on the data used in the construction of the RDI see Appendix D.4. An important caveat is that the diversity captured by the index could be diversity across different communities within a county rather than diversity within communities in the county.

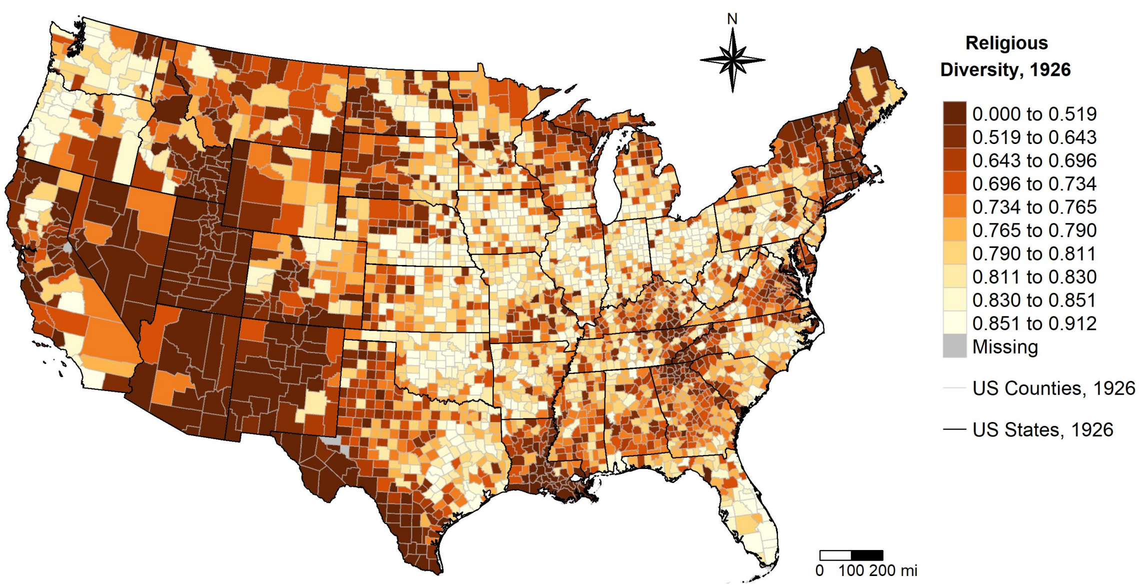
To get a sense of the variation in the data, I plot in Figure A.1 the county-level RDI in the contiguous U.S. in 1926. A darker color implies a lower RDI score. Note that the national spatial patterns that emerge from the map are somewhat different than the spatial pattern for the LNI (Figure 2) and the SFI (Figure B.1), which is particularly visible in New England. The within state pattern are more similar.

Results. I find a robust relationship between local soil heterogeneity and close-knit communities as measured by the RDI. Table A.1 reports estimates of equation 2 when the dependent variable is the RDI for four different specifications.[34] The specification in column 1 does not include any controls, and find no relationship between the SHI and the RDI. However once state-by-year fixed affects are included (column 2), the coefficient increases and the standard error decreases, so that the relationship becomes significant. When observable and unobservable geo-climatic characteristics are included (columns 3-4) the point estimate drops in magnitude, but remains highly significant. My preferred specification in column 4 suggests that an increase from a complete soil homogeneity (SHI = 0) to a complete soil

[34]This table is the analog of Table 2, which reports results when the dependent variable is the LNI.

heterogeneity (SHI = 1) is associated with an increase of 0.416 ($p-value$ = 0.001) standard deviations in the RDI. This result is robust to using alternative ways to account for spatial auto-correlation in the data (Appendix Figure C.3) and alternative distances in the calculation of the SHI (Appendix Table C.5). In column 5 I also control for birth place diversity. While birth place diversity seems to be positively correlated with religious diversity, as expected, controlling for it does not have a significant impact on the relationship between soil heterogeneity and religious diversity. Similar to the results in section 3.4.4, this suggests that an impact on the diversity of settlers' origin in not a key driver of the association between soil heterogeneity and close-knit communities. Those results provide further support to Prediction 1.

FIGURE A.1: RELIGIOUS DIVERSITY INDEX, 1926



Note: This figure plots the county-level *"Religious Diversity Index"* (RDI) in the contiguous U.S. in 1926. Darker color implies lower RDI. Data is from Manson et al., 2020 See Appendix D.4 for a description of the RDI construction.

TABLE A.1: SOIL HETEROGENEITY INCREASES RELIGIOUS DIVERSITY

	Dependent variable: Religious Diversity				
	(1)	(2)	(3)	(4)	(5)
Soil Heterogeneity	0.275	0.754***	0.457***	0.416***	0.393***
	(0.217)	(0.144)	(0.126)	(0.129)	(0.129)
Birth Place Diversity					0.113***
					(0.026)
Oster δ for $\beta = 0$		-3.34	-11.71	-19.58	-29.50
Number of Observations	19,881	19,881	19,881	19,881	19,868
Number of Counties	3,317	3,317	3,317	3,317	3,304
Number of Clusters	338	338	338	338	338
R^2	0.00093	0.35	0.39	0.4	0.41
State-by-Year Fixed Effects		✓	✓	✓	✓
Geoclimatic Controls			✓	✓	✓
Smooth Location Controls				✓	✓

Note: This table reports estimates of equation 2 when the dependent variable is the RDI. The data is from Manson et al., 2020. See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p <$ 0.10, ** $p < 0.05$, *** $p < 0.01$

B The Strength of Family Ties

This section studies the relationship between local soil heterogeneity and the *"Strength of Family Ties"* (SFT). I focus on the SFT as an alternative measure for close-knit social networks for two main reasons. First, research in social psychology had identified family ties as a key factor that correlates with interdependence across cultures (Triandis et al., 1990; Triandis, 2001; Vandello and Cohen, 1999).[35] In the economics literature, strong family ties have been shown to correlate with many personal and cultural feature, that in turn were associated with close-knit social networks and interdependence more broadly, such as lower geographical mobility, generalized trust, and adverse attitudes toward changes (Alesina and Giuliano, 2010, 2011, 2014; Alesina et al., 2015). The second reason is simply that, much like naming patterns, information on family structure and the choice of living arrangements is observable in historical census data.

I use data on family structure and the choice of living arrangements from the full count census data between 1860-1940 to construct a time-varying county-level SFT measure.[36] For a description of the SFT construction see Appendix D.3.

To get a sense of the variation in the data, I plot in Figure B.1 the county-level SFT in the contiguous U.S. in 1940. A darker color implies a higher SFT score. The spatial cultural patterns that emerge from the map resemble the SHI spatial patterns (see Figure 1). The Northeast, the "rust belt" and most of the West tend to have low SFT, while the South, the "wheat belt" and Utah tend to have high SFT. Within states, counties that are home to large cities (e.g., Fulton County, Georgia, where Atlanta is located) tend to have a lower SFT compared to other counties in the same state.

Results. I find a robust relationship between local soil heterogeneity and interdependence as measured by the SFT. Table B.1 reports estimates of equation 2 when the dependent variable is the SFT for four different specifications.[37] Across all specifications, I find that soil heterogeneity decreases the SFT, thereby providing further validity to the empirical results supporting Prediction 1 that I reported in section 3.3. Depending on the controls included in the regression, I find that an increase from a complete soil homogeneity (SHI = 0) to a complete soil heterogeneity (SHI = 1) is associated with a decrease of between 0.380 and 0.451 standard deviations in the SFT.

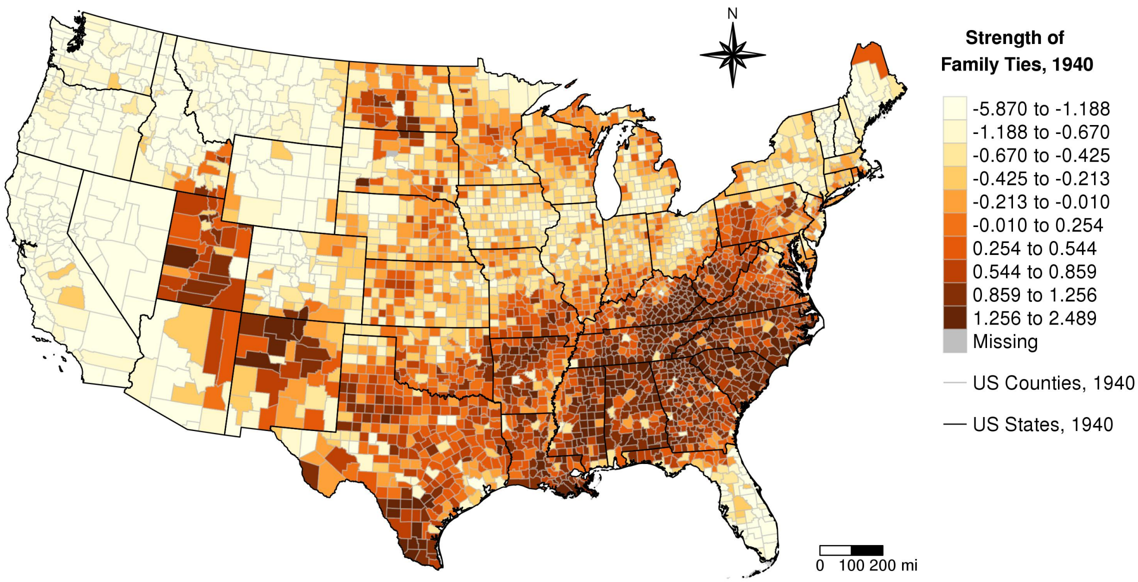
This result is robust to using alternative ways to account for spatial auto-correlation in the data (Appendix Figure C.4), and alternative distances in the calculation of the SHI (Appendix Table C.6).

[35] The social psychology literature often use the term "family integrity."

[36] 1850 is excluded because in this year information regarding material status was not recorded.

[37] This table is the analog of Table 2, which reports results when the dependent variable is the LNI.

FIGURE B.1: STRENGTH OF FAMILY TIES, 1940



Note: This figure plots the county-level *"Strength of Family Ties"* (SFT) in the contiguous U.S. in 1940. Data includes all individuals not living in group quarters in the 1940 Census. Darker color implies higher SFT. See Appendix D.3 for a description of the SFT construction.

TABLE B.1: SOIL HETEROGENEITY REDUCES THE STRENGTH OF FAMILY TIES

	Dependent variable: Strength of Family Ties			
	(1)	(2)	(3)	(4)
Soil Heterogeneity	-1.323^{***}	-0.357^{**}	-0.424^{**}	-0.444^{**}
	(0.299)	(0.181)	(0.170)	(0.178)
Oster δ for $\beta = 0$		0.74	0.97	1.18
Number of Observations	21,736	21,736	21,736	21,736
Number of Counties	3,460	3,460	3,460	3,460
Number of Clusters	338	338	338	338
R^2	0.021	0.55	0.57	0.59
State-by-Year Fixed Effects		✓	✓	✓
Geoclimatic Controls			✓	✓
Smooth Location Controls				✓

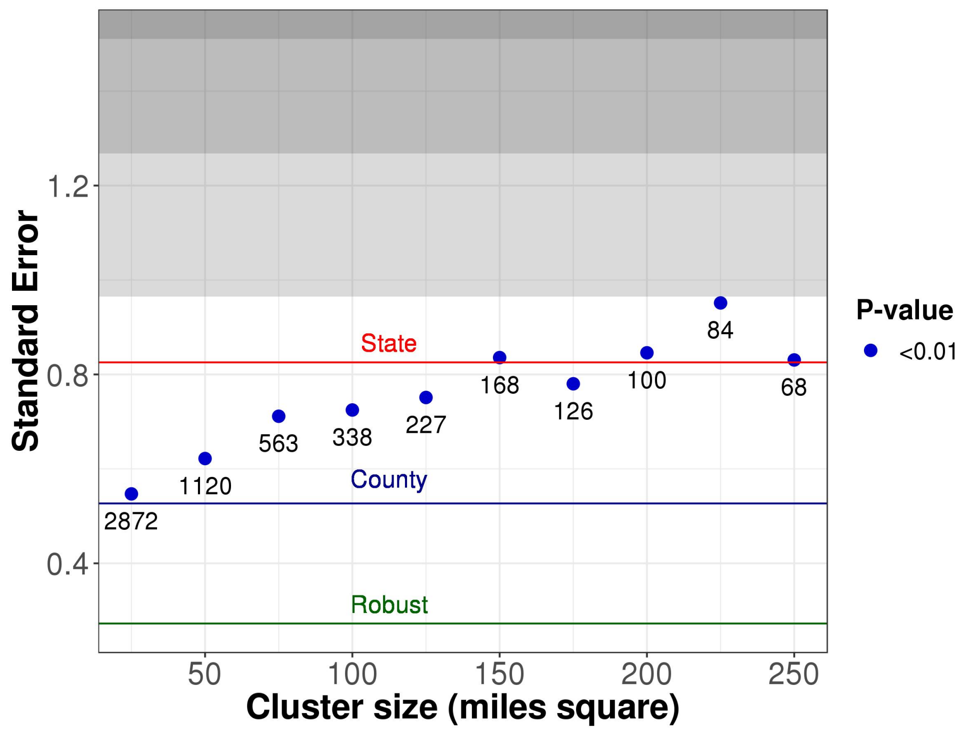
Note: This table reports estimates of equation 2 when the dependent variable is the baseline county-level SFT. The data is from the full count censuses between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C Robustness Checks

C.1 Robustness of the Main Result

This Appendix section presents several robustness checks with the results discussed in section 3.3. Appendix Figure C.1 documents robustness to alternative ways to account for spatial auto-correlation, Appendix Table C.1 documents robustness to alternative definitions of the LNI, and Appendix Table C.2 documents robustness to alternative definitions of the SHI.

FIGURE C.1: INFERENCE ROBUSTNESS



Note: This figure plots the standard error of β from the preferred specification of equation 2 when the dependent variable is the local name index (LNI), using different approaches for inference. The blue points represent the standard errors (on the y-axis) using arbitrary grid-cell of different sizes (on the x-axis), as proposed by Bester et al. (2011). The numeric label under each point indicates the number of spatial clusters. The green horizontal line plots the HC robust standard errors, the dark blue horizontal line plots the standard errors when clustering at the county level, and the red horizontal line plots the standard errors when clustering at the state level. The background color is indicative of the level of statistical significance. The p-value is < 0.01 in the white area, and < 0.05, < 0.1 and > 0.1 in the light to dark shades of gray. See Appendix D for details on the data and variable construction.

TABLE C.1: ALTERNATIVE DEFINITIONS OF THE LNI

	Dependent variable: Local Name Index					
LNI definition	Baseline	Include foreign-born parents	Include all races	Include all races and foreign-born parents	At least 100 name repetitions	State defined as local
	(1)	(2)	(3)	(4)	(5)	(6)
Soil Heterogeneity	-2.486^{***}	-2.836^{***}	-2.458^{***}	-2.839^{***}	-2.416^{***}	-2.197^{***}
	(0.725)	(0.723)	(0.704)	(0.709)	(0.721)	(0.746)
Number of Observations	23,435	23,461	23,453	23,471	23,433	23,435
Number of Counties	3,499	3,506	3,501	3,508	3,499	3,499
Number of Clusters	338	338	338	338	338	338
R^2	0.55	0.55	0.55	0.54	0.53	0.43
Dependent Variable Mean	67.82	67.39	67.76	67.28	64.50	57.11
Dependent Variable SD	6.30	6.15	6.02	5.87	6.54	4.30
State-by-Year Fixed Effects	✓	✓	✓	✓	✓	✓
Geoclimatic Controls	✓	✓	✓	✓	✓	✓
Smooth Location Controls	✓	✓	✓	✓	✓	✓

Note: This table reports estimates of equation 2 when the dependent variable is children's LNI under different definitions. "Local" is defined as the county in columns 1-5 and as the state in column 6. The base sample include white native-born children between the age of 0 to 10 with native-born parents in columns 1, 5, and 6. In column 2 the sample also includes children of foreign-born parents, in column 3 it also includes non-white children, and in column 4 it includes all native-born children between the age of 0 to 10. Column 5 further restricts the sample to include names that are observed at least 100 time nationally within the same year. The data is from the full count censuses between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE C.2: ALTERNATIVE DISTANCE FOR SHI CALCULATION

	Dependent variable: Local Name Index						
SHI Cell Distance	Baseline 25	10	15	20	30	35	40
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Soil Heterogeneity	-2.486^{***}	-2.210^{***}	-2.329^{***}	-2.423^{***}	-2.535^{***}	-2.575^{***}	-2.612^{***}
	(0.725)	(0.788)	(0.744)	(0.726)	(0.732)	(0.742)	(0.755)
Number of Observations	23,435	23,435	23,435	23,435	23,435	23,435	23,435
Number of Counties	3,499	3,499	3,499	3,499	3,499	3,499	3,499
Number of Clusters	338	338	338	338	338	338	338
R^2	0.55	0.55	0.55	0.55	0.55	0.55	0.55
Dependent Variable Mean	67.82	67.82	67.82	67.82	67.82	67.82	67.82
Dependent Variable SD	6.30	6.30	6.30	6.30	6.30	6.30	6.30
State-by-Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Geoclimatic Controls	✓	✓	✓	✓	✓	✓	✓
Smooth Location Controls	✓	✓	✓	✓	✓	✓	✓

Note: This table reports estimates of equation 2 when the dependent variable is children's LNI in which "local" is defined as the county, when the SHI is calculated over different areas (cell distances). The data is from the full count censuses between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.2 Robustness of Difference-in-Differences and Triple-Difference Results

TABLE C.3: ADDING INDIVIDUAL CONTROLS TO DID AND TRIPLE-D ESTIMATES

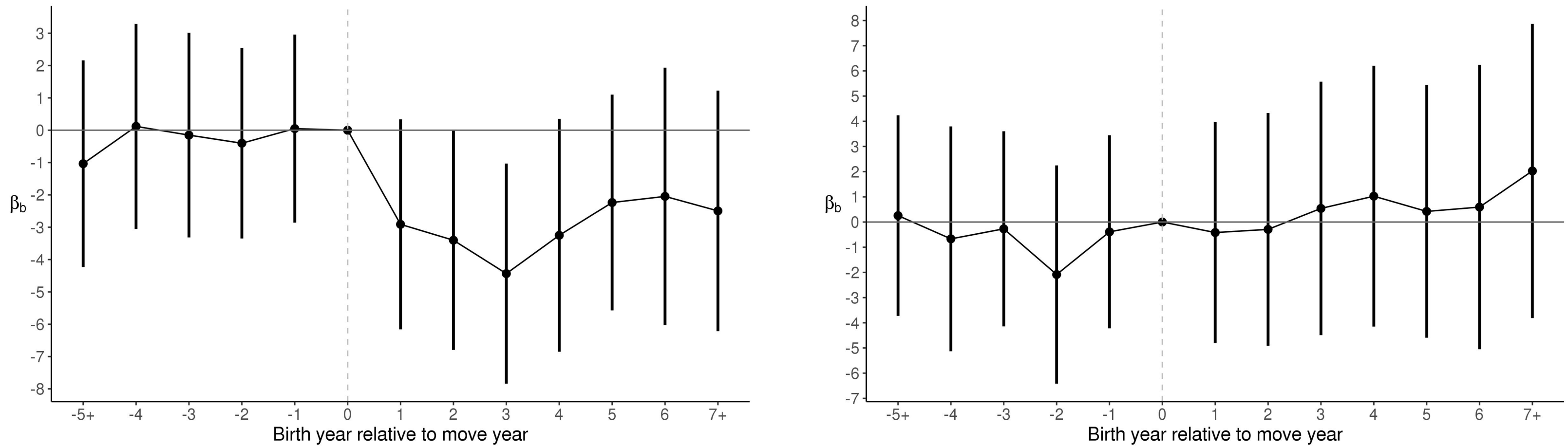
	Dependent variable: Local Name Index			
	Difference-in-Differences			Triple-Difference
Sample:	All Households	Farmer′s Households	Non-Farmer′s Households	All Households
	(1)	(2)	(3)	(4)
Post Move × Soil Heterogeneity	−1.709**	−3.054***	0.791	0.844
	(0.824)	(0.865)	(1.170)	(1.152)
Post Move × Farmers′ Household × Soil Heterogeneity				−3.962***
				(1.179)
Number of Observations	1,203,908	713,881	490,022	1,203,903
Number of Counties	2,559	2,472	2,477	2,559
R^2	0.37	0.35	0.39	0.37
Dependent Variable Mean	54.22	54.40	53.95	54.22
Dependent Variable SD	13.66	13.42	14.00	13.66
Households Fixed Effects	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓

Note: This table presents estimates of equations 5 and 7 when the dependent variable is children's LNI in which "local" is defined as the state. The regressions also control for for child's gender, birth order and a 5-year cohort fixed effect. The data is from the full count censuses between 1850-1880 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered by county in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

FIGURE C.2: ADDING INDIVIDUAL CONTROLS TO DID ESTIMATES

(A) Difference-in-Differences: Farmers' Families

(B) Difference-in-Differences: Non-Farmers' Families



Note: This figure plots the estimates of β_b and 95% confidence intervals from equations 4 and 6 when the dependent variable is children's LNI where "local" is defined as the state. The regressions also control for for child's gender, birth order and a 5-year cohort fixed effect. The data is from the full count censuses between 1850-1880 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at the county in parentheses.

TABLE C.4: ROBUSTNESS OF DIFFERENCE-IN-DIFFERENCES ESTIMATES TO SAMPLE SELECTION

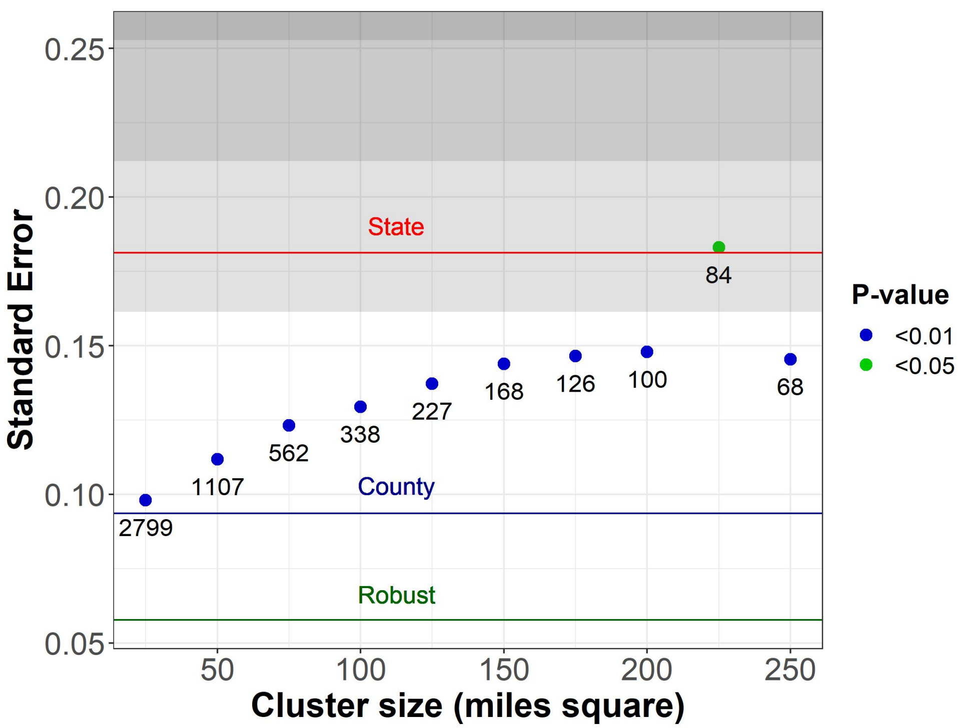
	Dependent variable: Local Name Index		
Sample:	All Households	Farmer′s Households	Non-Farmer′s Households
	(1)	(2)	(3)
	Panel A: Include foreign-born parents		
Post Move × Soil Heterogeneity	-2.150^{***}	-2.802^{***}	-0.991
	(0.800)	(0.837)	(1.352)
	Panel B: Include all races		
Post Move × Soil Heterogeneity	-2.300^{***}	-3.673^{***}	0.081
	(0.851)	(0.931)	(1.112)
	Panel C: Include multiple moves		
Post Move × Soil Heterogeneity	-1.738^{**}	-3.195^{***}	0.898
	(0.825)	(0.875)	(1.152)
Households Fixed Effects	✓	✓	✓

Note: This table presents estimates of equation 5 when the dependent variable is children's LNI in which "local" is defined as the state. The data is from the full count censuses between 1850-1880 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered by county in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.3 Robustness of Religious Diversity Result

This Appendix section presents several robustness checks for the result on the RDI. Appendix Figure C.3 documents robustness to alternative ways to account for spatial auto-correlation, and Appendix Table C.5 documents robustness to alternative definitions of the SHI.

FIGURE C.3: RDI. INFERENCE ROBUSTNESS



Note: This figure plots the standard error of β from the baseline specification of equation 2 when the dependent variable is the religious diversity index (RDI), using different approaches for inference. The blue points represent the standard errors (on the y-axis) using arbitrary grid-cell of different sizes (on the x-axis), as proposed by Bester et al. (2011). The numeric label under each point indicates the number of spatial clusters. The green horizontal line plots the HC robust standard errors, the dark blue horizontal line plots the standard errors when clustering at the county level, and the red horizontal line plots the standard errors when clustering at the state level. The background color is indicative of the level of statistical significance. The p-value is < 0.01 in the white area, and < 0.05, < 0.1 and > 0.1 in the light to dark shades of gray. See Appendix D for details on the data and variable construction.

TABLE C.5: RDI. ALTERNATIVE DISTANCE FOR SHI CALCULATION

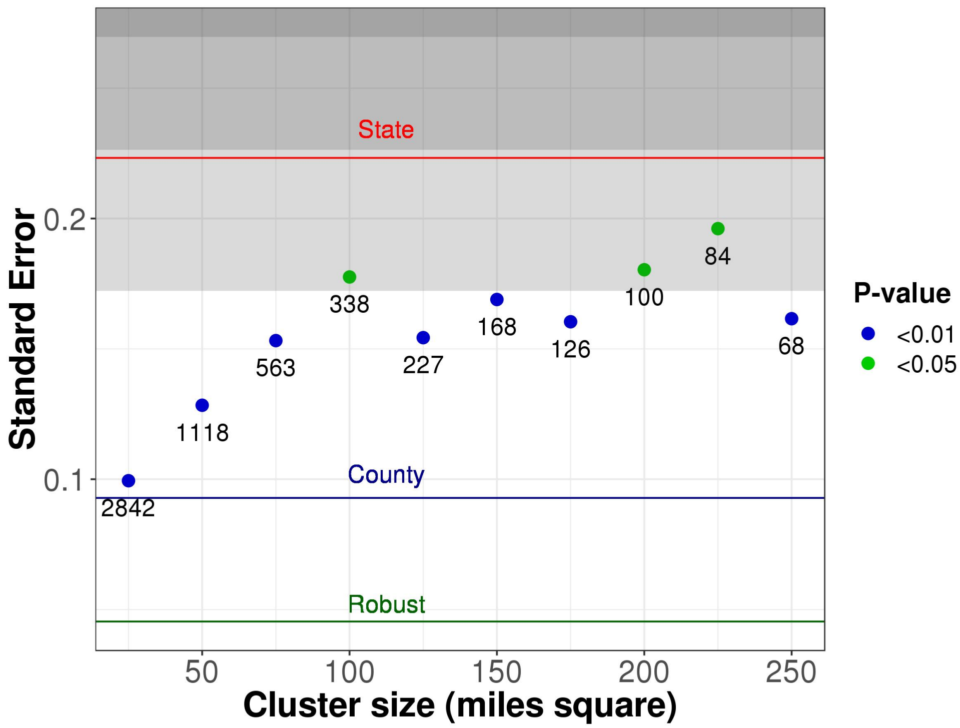
	Dependent variable: Religious Diveristy Index						
SHI Cell Distance	Baseline 25	10	15	20	30	35	40
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Soil Heterogeneity	0.416***	0.387***	0.404***	0.413***	0.415***	0.411***	0.406***
	(0.129)	(0.135)	(0.129)	(0.128)	(0.132)	(0.134)	(0.137)
Number of Observations	19,881	19,881	19,881	19,881	19,881	19,881	19,881
Number of Counties	3,317	3,317	3,317	3,317	3,317	3,317	3,317
Number of Clusters	338	338	338	338	338	338	338
R^2	0.4	0.4	0.4	0.4	0.4	0.4	0.4
State-by-Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Geoclimatic Controls	✓	✓	✓	✓	✓	✓	✓
Smooth Location Controls	✓	✓	✓	✓	✓	✓	✓

Note: This table reports estimates of equation 2 when the dependent variable is the RDI, when the SHI is calculated over different areas (cell distances).See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.4 Robustness of Strength of Family Ties Result

This Appendix section presents several robustness checks for the result on the SFT. Appendix Figure C.4 documents robustness to alternative ways to account for spatial auto-correlation, and Appendix Table C.6 documents robustness to alternative definitions of the SHI.

FIGURE C.4: SFT. INFERENCE ROBUSTNESS



Note: This figure plots the standard error of β from the baseline specification of equation 2 when the dependent variable is the strength of family ties (SFT), using different approaches for inference. The blue points represent the standard errors (on the y-axis) using arbitrary grid-cell of different sizes (on the x-axis), as proposed by Bester et al. (2011). The numeric label under each point indicates the number of spatial clusters. The green horizontal line plots the HC robust standard errors, the dark blue horizontal line plots the standard errors when clustering at the county level, and the red horizontal line plots the standard errors when clustering at the state level. The background color is indicative of the level of statistical significance. The p-value is < 0.01 in the white area, and < 0.05, < 0.1 and > 0.1 in the light to dark shades of gray. The data is from the full count censuses between 1860-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction.

TABLE C.6: SFT. ALTERNATIVE DISTANCE FOR SHI CALCULATION

	Dependent variable: Strength of Family Ties						
SHI Cell Distance	Baseline 25	10	15	20	30	35	40
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Soil Heterogeneity	-0.444^{**}	-0.275	-0.354^{*}	-0.407^{**}	-0.468^{***}	-0.485^{***}	-0.499^{***}
	(0.178)	(0.186)	(0.182)	(0.179)	(0.176)	(0.176)	(0.175)
Number of Observations	21,736	21,736	21,736	21,736	21,736	21,736	21,736
Number of Counties	3,460	3,460	3,460	3,460	3,460	3,460	3,460
Number of Clusters	338	338	338	338	338	338	338
R^2	0.59	0.59	0.59	0.59	0.59	0.59	0.59
State-by-Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Geoclimatic Controls	✓	✓	✓	✓	✓	✓	✓
Smooth Location Controls	✓	✓	✓	✓	✓	✓	✓

Note: This table reports estimates of equation 2 when the dependent variable is the SFT, when the SHI is calculated over different areas (cell distances). The data is from the full count censuses between 1860-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). See Appendix D for details on the data and variable construction. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D Data Sources and Variables Construction

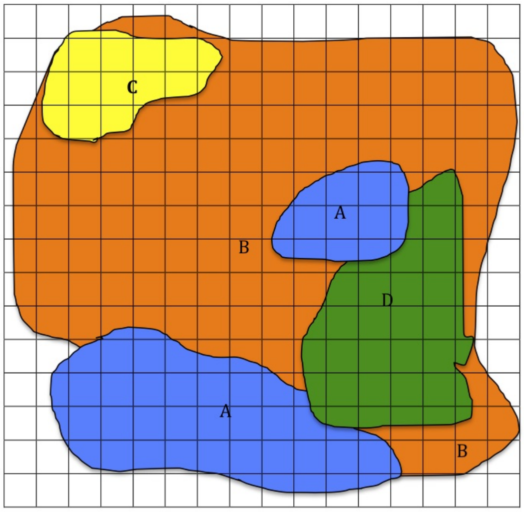
D.1 Soil Heterogeneity Index

I use detailed geo-referenced soil data from the Digital General Soil Map of the United States (STATSGO2) (Soil Survey Staff, 2017b), which maps soil areas that can be shown at a map of scale 1:250,000 in the continental U.S., to construct a county-level *"Soil Heterogeneity Index"* (SHI). The index is meant to capture the average dissimilarity of soil across neighboring farmers. I construct the SHI in the following steps:

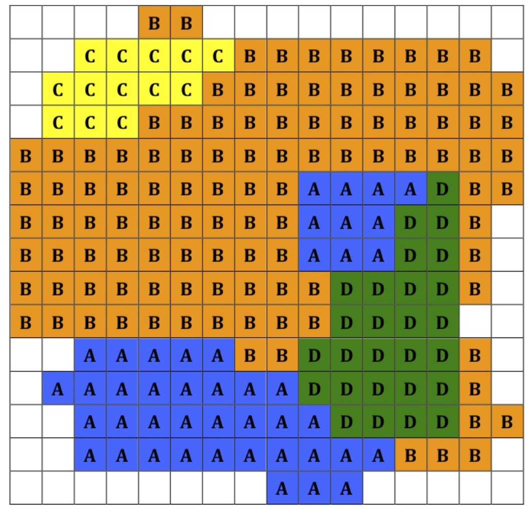
1. I convert the STATSGO2 map containing polygon features into a raster dataset containing fine-grid cells of size 500 meters square (illustrated in Figure D.1).
2. For each cell, I calculate the probability that a randomly selected neighboring cell is of a different soil type (illustrated in Figures D.2-D.3).

   I define neighboring cells as cells that fall within a square of a given size around each cell ("the considered area") rather than in terms of actual distance (i.e. fall within a given circle around each cell) to reduce computational demand. The size of the considered area affects the degree of heterogeneity. The bigger the area, the more likely it is to have cells of different soil types. My baseline SHI uses half of the mean size of U.S. counties in 2000 as a benchmark for the size of the considered area. I document robustness to using different sizes of the considered area.
3. Last, I aggregate the SHI at the fine grid level to the county level by taking the mean grid-level SHI within the county.

Figure D.1: SHI Construction. STATSGO2 Data to Grid-Cells



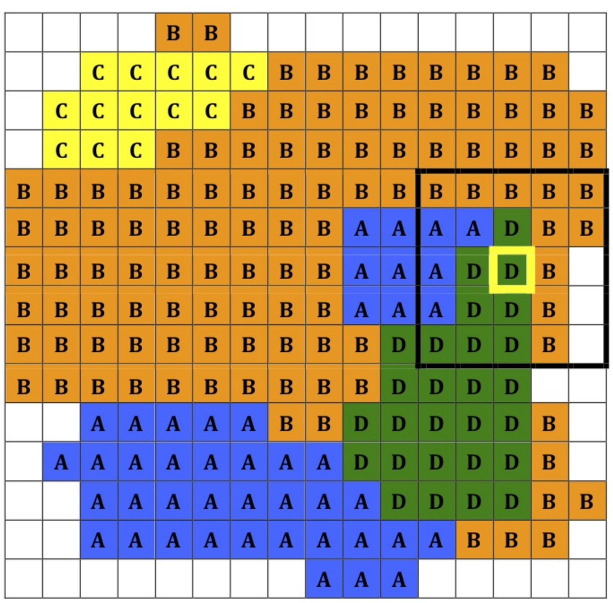
(A) Underlying STATSGO2 Data

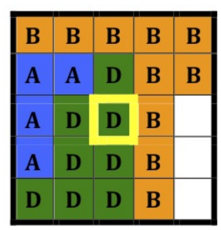


(B) Grid-cells data

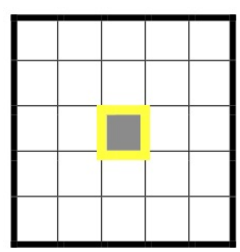
Note: This figure illustrates step (1.) of the county-level SHI construction - converting the STATSGO2 map containing polygon features (Figure A) into a raster dataset containing fine-grid cells of size 500 meters square (Figure B).

Figure D.2: SHI Construction. Calculating SHI for each Cell





(B) 7 same-type cells out of 21 neighboring soil cells

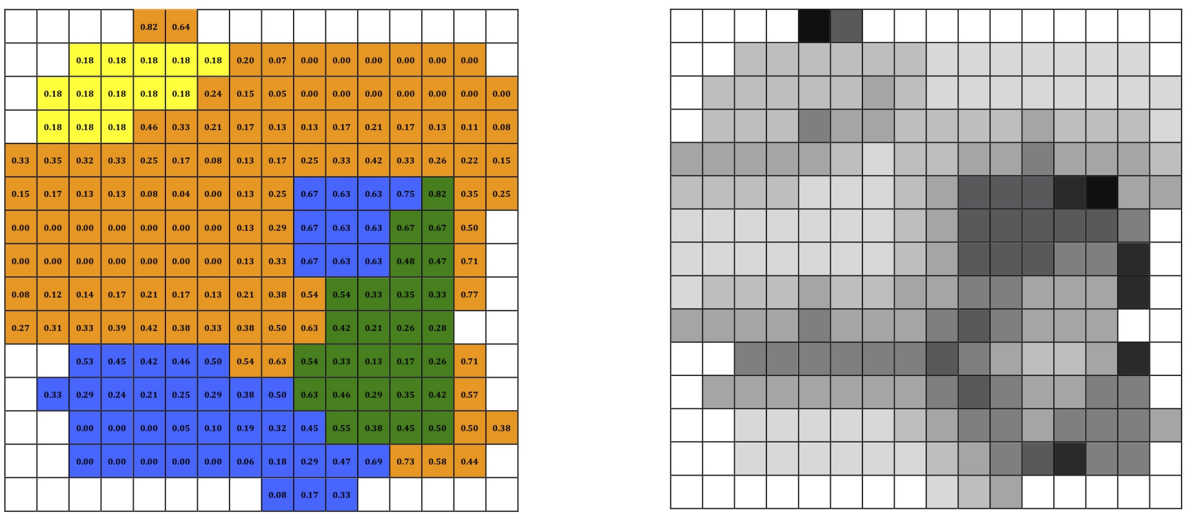


(A) Neighboring cells (in black) around a given cell (in yellow)

(C) SHI for given cell is $1-7/24=0.67$

Note: This figure illustrates step (2.) of the county-level SHI construction - calculating the the probability that a randomly selected neighboring cell (in bold-black frame) is of a different soil type than a given cell (in bold-yellow frame).

FIGURE D.3: SHI CONSTRUCTION. SHI FOR EACH CELL



(A) Numeric representation

(B) Heat map representation

Note: This figure illustrates the result of applying step (2.) of the county-level SHI construction to each cell. Figure (A) represents the cell-level SHI numerically, and Figure (B) represents it as a heat map.

D.2 Local Name Index

Definition. The LNI is defined as

$$LNI_{first\ name,l,g,t} = 100 \times \frac{Pr(first\ name|l,g,t)}{Pr(first\ name|l,s,t) + Pr(first\ name|-l,g,t)} \tag{8}$$

where l is the geographical level defined as "local" - the contemporaneous county or state, g is the child's gender, and t is the census year. The index ranges from 0 to 100, where a value of 100 reflects a distinctively local name and a value of zero reflects a distinctively "outsider's" name. Note that the LNI is invariant to the size of the population in different localities and to the general popularity of a given name.

Data. I use data on children's first names from the full count censuses between 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019). The sample is restricted to only includes native-born children between the ages of 0 to 10. The baseline sample is further restricted to only includes white children with native-born parents. Alternative samples add to the baseline sample (i) children with foreign-born parents (i) non-white children, and (iii) both.

D.3 The Strength of Family Ties

I use data on family structure and the choice of living arrangements from the full count censuses between 1860-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019) to construct a county-level measure of the *"Strength of Family Ties"* (SFT).[38]

For each county-year, I calculate (i) the divorce-to-marriage ratio, (ii) the share of elderly people living without a relative, (iii) the share of people living with at least one person who is not their relative, and (iv) the mean size of families. Then, for each year, I conduct a principal component analysis at the county-level using these variables as inputs. The first eigenvector, which I refer to as the SFT, explains between $54-68\%$ of the variance in the four variables, depending on the year. It is also the only component with an eigenvalue that is larger than one in all years ($2.15-2.73$, depending on the year).[39] In all years the loading on the four variables always have the same sign (negative on divorce-to-marriage ration, the share of elderly people living without a relative and the share of people living with a non-relative, and positive on family size). Because there is no natural interpretation for the SFT units, I standardize it into z-scores within each year to ease the interpretation of estimated

[38] 1850 is excluded because in this year information regarding material status was not recorded.

[39] In 1860 the second component had an eigenvalue of 1.00. In all other years all other components are strongly smaller than 1.

effects.

D.4 Religious Diversity Index

Definition. The RDI is defined as

$$Religious\ Diversity\ Index_{ct} = 1 - \sum_j s_{cjt}^2$$

where s_{cjt} is the share of members of religious denomination j, in county c, and in year t, out of the total number of members in religious institutions in county c year t. Note that the RDI equals one minus the Herfindahl–Hirschman Index over the share of members of religious denominations. It measures the probability that two randomly drawn individuals from the population of members of religious institutions in a county belong to a different denomination.

The list of religious denominations for which data is collected varies across years. Therefore, and to ease the interpretation of estimated effects, in the empirical analysis I standardize the RDI into z-scores within each year.

Data. To calculate the index, I use county-level data on the number of members of religious institutions by denomination between for the years 1850, 1860, 1870, 1890, 1906, 1916, and 1926 (Manson et al., 2020).

D.5 Agricultural Diversity

Definition. Agricultural diversity is defined as

$$Agricultural\ Diversity_{ct} = 1 - \sum_j s_{cjt}^2$$

where s_{cjt} is the share of acres used in the cultivation of agricultural product j, in county c, and in year t, out of the total number of acres under cultivation in county c year t. Note that the index equals one minus the Herfindahl–Hirschman Index over the share of acres used in the cultivation of different agricultural products. The agricultural diversity index measures the probability that two randomly drawn acres used in farms in a county are used to grow different agricultural products. The list of agricultural products for which data is collected varies across years. Therefore, and to ease the interpretation of estimated effects, in the empirical analysis I standardize the index into z-scores within each year.

Data. To calculate the index, I use county-level data on the number of acres used in the production of different agricultural products for the each decade between years 1880-1930, as well as 1925 and 1935 (Manson et al., 2020).

D.6 Other variables

LNI validation.

Relative importance of communal values. One minus the county-level variable named "universalist_vs_communal_values" which measures the relative importance of universalist vs. communal moral values and taken from Enke (2020). Data is available on-line in the authors homepage. For details on the construction of the variable, see Enke (2020).

Trump vote share 2016. County-level data on Trump's vote share in the 2016 presidential election, standardized into z-scores. Data from Leip (2017). The use if this variable to validate the LNI measure is inspired by Enke (2020).

Δ *[Trump - GOP].* County-level data on the difference between Trump's vote share in the 2016 presidential election and the average vote share of Republican presidential candidates in the 2000-2012 presidential elections, standardized into z-scores. Data from Leip (2017). The use if this variable to validate the LNI measure is inspired by Enke (2020).

Geo-climatic controls.

Average temperature. County-level mean annual temperature (in Celsius degrees), calculated for all contemporary U.S. counties in each decade using GIS and information on counties' contemporary borders from Manson et al., 2020 and a 5 arc-minutes mean annual temperature raster (baseline period, 1961-1990) from IIASA/FAO (2012).

Average precipitation. County-level mean annual precipitation (in mm), calculated for all contemporary U.S. counties in each decade using GIS and information on counties' contemporary borders from Manson et al., 2020 and a 5 arc-minutes mean annual precipitation raster (baseline period, 1961-1990) from IIASA/FAO (2012).

Average elevation. County-level mean elevation (in meters), calculated for all contemporary U.S. counties in each decade using GIS and information on counties' contemporary borders from Manson et al., 2020 and a 30 arc-second elevation raster (voild-filled DEM) from Lehner et al. (2008).

Average slope. County-level mean slope (in radians), calculated for all contemporary U.S. counties in each decade using GIS and information on counties' contemporary borders from Manson et al., 2020 and a 30 arc-second elevation raster (voild-filled DEM) from Lehner et al. (2008).

Flow accumulation. County-level mean flow accumulation, defined as the amount of upstream area (in number of cells) draining into each cell and measuring the measure of the upstream catchment area, calculated for all contemporary U.S. counties in each decade using GIS and information on counties' contemporary borders from Manson et al., 2020 and a 30 arc-second flow accumulation raster (voild-filled ACC) from Lehner et al. (2008).

River density. County-level river density, defined as share of county area in rivers or streams assuming a fixed 10-meters width for all rivers and streams, calculated for all contemporary U.S. counties in each decade using GIS and information on counties' contemporary borders from Manson et al., 2020 and a 30 arc-second river network raster (voild-filled RIV) from Lehner et al. (2008).

Average absolute agricultural productivity. County-level average maximal potentially dollar value of agricultural products in 1860 prices. Calculated for the following crops: barley, buckwheat, cotton, flax, maize, oat, rye, sugercane, sweet potato, tobacco, rice, wheat, and white potato, using potential production capacity rasters (in t/ha), under intermediate input level and rain-fed management, baseline period, 1961-1990 from IIASA/FAO (2012), and data on the price of farm products 1859/1860 from Manson et al., 2020.

Farmers.

Share farmers. The share of households in a county in the relevant sample with an occupation code (1950 basis) of 100- "Farmers (owners and tenants)," or 123 - "Farm managers," out of the total number of households in sample in the county with a valid occupation code.

Farmers' household. A binary variable that equals one if the father's occupation code (1950 basis) is either 100- "Farmers (owners and tenants)," or 123 - "Farm managers," and zero if the father has another valid occupation code.

Fertilizers.

Share Using Fertilizers. County level data on the share of farms reporting expenditure on fertilizers out of the total number of farms for the years 1910-1930 (Manson et al., 2020).

Growth in Fertilizers Use. The growth in the share of farms using fertilizers from the previous period.

Confounding Channels.

Farms' size Gini. The Gini coefficient on the county-level distribution of farm sizes for the years 1860-1940 (Manson et al., 2020). The coefficient is calculated using the midpoint of each category and 125% of the last (unbounded) category.

Birth Place Diversity. A county-level index that equals one minus the Herfindahl–Hirschman Index over the share of household heads that were born in different countries (or regions) for the years 1850-1940 (Ruggles et al., 2020; Minnesota Population Center, 2019).

Moral Foundations Questionnaire (MFQ).

Data on individual responses to the MFQ (Graham et al., 2011), which was developed to measure the degree to which individuals' moral judgment involves the five foundations highlighted by the *"Moral Foundations Theory"* (MFT) (Haidt and Graham, 2007). Data surveyed on www.yourmorals.org between 2008-2018, and includes the individual responses of approximately $242{,}000$ Americans. For more information on the MFT and its measure, including the full MFQ texts, see https://moralfoundations.org/. Moral foundations scores range from 0 to 30, then standardized into z-scores.

Harm / Care. Measures the importance of virtues such as kindness, gentleness, and nurturance.

Fairness / Reciprocity. Measures the importance of virtues such as justice, rights, and autonomy.

In-group / Loyalty. Measures the importance of virtues such as patriotism and self-sacrifice for the group.

Authority / Respect. Measures the importance of virtues such as leadership and followership, including deference to legitimate authority and respect for traditions.

Purity / Sanctity. Measures the importance of virtues such as striving to live in an elevated, less carnal, more noble way.

Binding versus Individualizing. The first eigenvector from a principal component analysis on the five foundations above. It explains 46% of the variance in the five foundations and has an eigenvalue of 2.29. It is the only eigenvector for which the signs of the loadings on the five foundations corresponds to the "binding" versus "individualizing" distinction (loads negatively on Harm / Care and Fairness / Reciprocity, and positively on In-group / Loyalty, Authority / Respect and Purity / Sanctity).