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

Itzchak Tzachi Raz

The Hebrew University of Jerusalem

NBER SI DAE
July 13, 2021

Question:

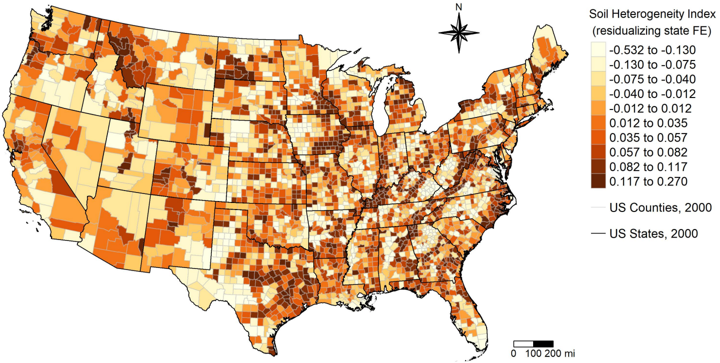
What happens to the strength of communal ties when beneficial social learning takes place within the community?

The settlement of the United States

- Millions of farmers migrated to new environments with unfamiliar characteristics (Shannon, 1945; Olmstead and Rhode, 2008)
- Need to quickly discover the optimal location-specific farming practices
- Two possible strategies:
  1. Learning by doing– individual trial and error
  2. Social learning– build on the experience of their neighbors
- However, substantial heterogeneity of soil in their area limited the effectiveness of social learning
- According to the historian Fred Shannon (1945), farmers' inability to rely on social learning fostered their "*traditional individualism*"

Example Testable predictions

County-level Soil Heterogeneity Index (SHI)



Note: This figure plots the soil heterogeneity index (SHI) for counties in the contiguous U.S. in 2000. Darker color implies a higher soil heterogeneity.

▸ Soil ▸ Details ▸ Raw data

Local Name Index (LNI)

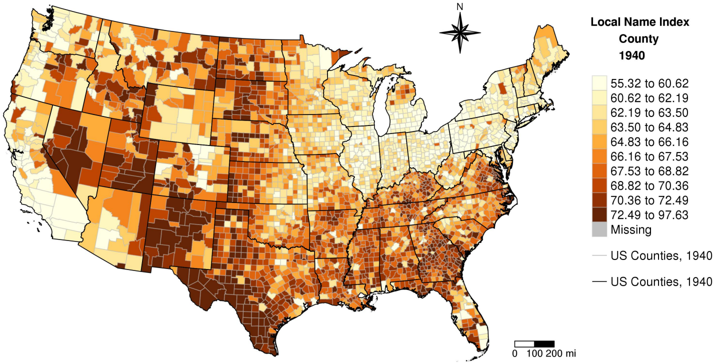
- I proxy close-knit communities with the centrality of communal identity in individuals' self-definition, and infer the later from children naming patterns ▸ More
- I follow Fryer and Levitt (2004) and construct a *"Local Name Index"* (LNI) using children first names from the full count censuses 1850-1940
- The probability that a name is given to a local child relative to a child in different locations in the U.S. ▸ Example

$$LNI_{nlgt} = 100 \times \frac{Pr(n|l, g, t)}{Pr(n|l, g, t) + Pr(n| - l, g, t)}$$

where n is a first name, l is the geographical level defined as "local," 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

- Alternative proxy measurements for a close-knit social structure:
  - Religious diversity and the strength of family ties

County-level Local Name Index (LNI), 1940



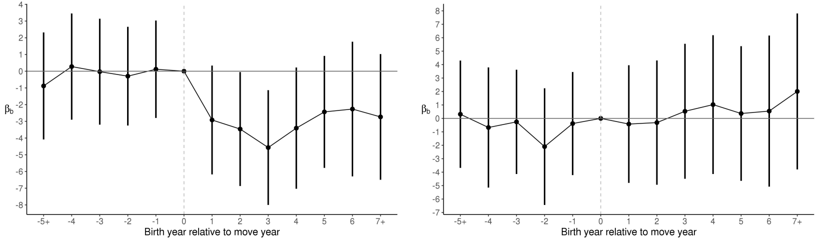
Note: This figure plots the county-level *"Local Name Index"* (LNI) in which "local" is defined as the county, for counties in the contiguous U.S. in 1940. Data includes native-born white children between the age of 0 to 10 with native-born parents in the 1940 Census. Darker color implies a higher local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

▸ Validation using Enke (2020)

Causal estimates: Difference-in-Differences

Soil heterogeneity only impacts farmers

- Study naming patterns of children born in families that migrated to counties with varying degrees of soil heterogeneity, before and after the family moved



Farmers' Families Non-Farmers' Families

Note: This figure plots the estimates of β_b and 95% confidence intervals from equation 3 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. Standard errors clustered at the county in parentheses.

▸ Estimation framework ▸ All families ▸ Triple-Difference ▸ Robustness ▸ Table ▸ OLS ▸ Long run

Farmers' social learning? Suggestive evidence

- Soil heterogeneity affects farmers themselves ‣ Table
  - Soil heterogeneity is associated with culture even holding the share of farmers fixed
  - ... but it does not matter for culture if no one is engaged in agriculture
  - ... and it matters more the higher the share of farmers
- Soil heterogeneity is associated with higher agricultural diversity ‣ Table
- Soil heterogeneity is associated with a lower rate of fertilizers adoption, consistent with limiting farmer's social learning ‣ Table
- Two confounding channels - agricultural inequality and birthplace diversity, are unlikely to explain the association ‣ Table

Concluding Remarks

- The distinction between close-knit and loose-knit cultures is often considered as the fundamental cultural cleavage across societies
- I provide the first 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

Contributions to the literature

Thank you!

Appendix

“There were varied reason for this traditional individualism, and none was more potent than the farmer’s bafflement when confronted by natural forces. He was in perpetual conflict with climate he could not conquer and soils that he seldom understood. ***A local agricultural society, made up of individuals who were more prosperous, might declare that a given crop or specialty was best suited to the region. The less fortunate neighbor tries the same crop in an identical way; it failed, and he began to doubt. The most successful gentleman farmer of the area tried out a new fertilizer or method of tillage and declared it a success. Another followed his procedure and got worse crops than before.****”*

(Shannon, 1945, p. 4)

Back

“Social Learning Hypothesis”: Testable predictions

Prediction (Close-knit Communities)

All else equal, communities located in a higher soil heterogeneity environment are less likely to have close-knit social networks with interdependent members.

Prediction (Soil heterogeneity only impact farmers)

Soil heterogeneity should only directly impact the attributes and communal ties of farmers.

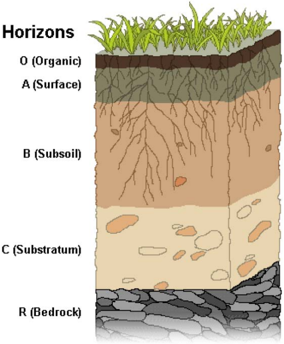
◂ Back ▸ Channels

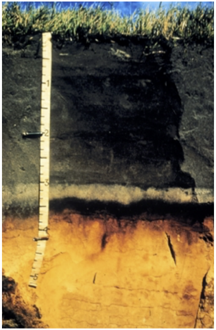
Soil

- 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
  - The nature of a soil varies across space
- Soil is characterized by distinguishable layers, formally referred to as “horizons”
  - The nature of a soil can not be determined from the surface alone
- **Soil taxonomy** is 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)
- The placement of a soil in soil taxonomy is broadly unaffected by cultivation

Back

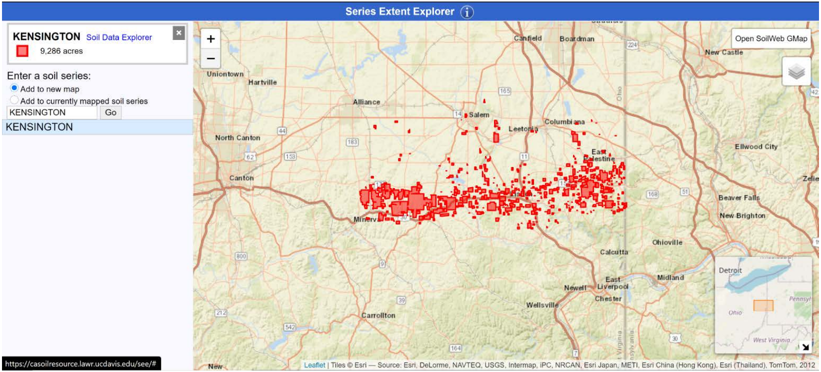
Soil horizons (Source: USDA “Soil Taxonomy”, 1999)





Back

Kensington soil series



Back

Kensington soil series

The Kensington series consists of deep, moderately well drained soils formed in loess, Illinoian age or early Wisconsinan age till, and residuum weathered from the underlying Pennsylvanian age shale, fine grained sandstone or siltstone on till plains. Permeability is moderate in the till and moderate or moderately rapid in the underlying material, above the bedrock. Slope ranges from 2 to 25 percent. Mean annual precipitation is about 37 inches, and mean annual temperature is about 51 degrees F.

TAXONOMIC CLASS: Fine-loamy, mixed, active, mesic Aquic Hapludults

Ap– 0 to 11 inches; brown (10YR 4/3) silt loam, pale brown (10YR 6/3) dry; weak fine granular structure; friable; common fine roots; 5 percent pebbles; moderately acid; abrupt smooth boundary. (7 to 11 inches thick.)

Bt1– 11 to 17 inches; strong brown (7.5YR 5/6) silt loam; weak medium subangular blocky structure; friable; common fine roots; few distinct brown (7.5YR 5/4) clay films on faces of peds; 10 percent medium prominent brown (10YR 5/3) clay depletions on faces of peds; 5 percent brown (10YR 4/3) Ap material in root and worm channels; 5 percent pebbles; very strongly acid; clear wavy boundary. (3 to 9 inches thick.)

⋮

USE AND VEGETATION: These soils are commonly used to grow corn, oats, wheat, mixed hay, and pasture. A few areas are in woodland. Native vegetation is mixed hardwood forest composed primarily of sugar maple, oak, and hickory.

Back

A "Soil Heterogeneity Index" (SHI)

- I use detailed geo-referenced soil data from the Digital General Soil Map of the United States (STATSGO2) to construct a county-level *"Soil Heterogeneity Index"* (SHI)

- The index ranges from 0 to 1 and captures the average dissimilarity of soil across neighboring locations in the county

- Calculated at a 500-m^2 cells level, the index equals the probability that a randomly selected neighboring cell is of a different soil type

- Then averaged within counties

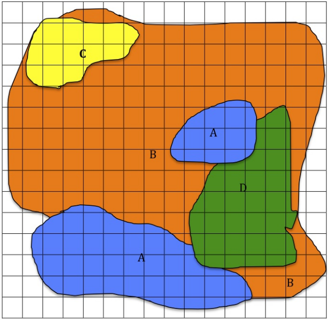
Back

SHI Construction

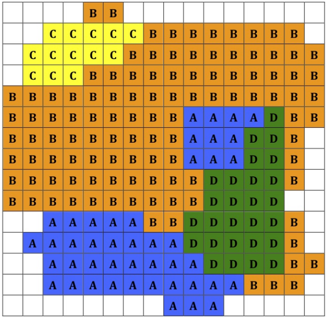
1. Convert the STATSGO2 map into a raster of cells of size 500 meters square
2. For each cell, calculate the probability that a randomly selected neighboring cell is of a different soil type
3. Last, I aggregate the SHI at the fine grid level to the county level by taking the mean grid-level SHI within the county.

Back

SHI Construction. STATSGO2 Data to Grid-Cells



(A) Underlying STATSGO2 Data

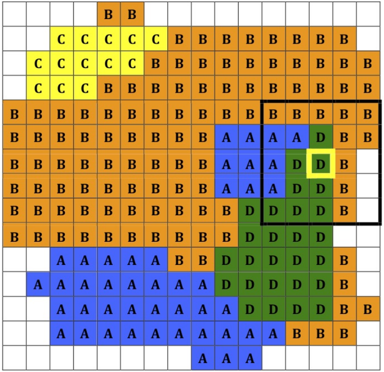


(B) Grid-cells data

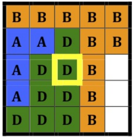
Note: This figure illustrates step (1.) of the county-level SHI construction - converting the STATSGO2 map containing polygon features (Figure A) into a raster dataset containing fine-grid cells of size 500 meters square (Figure B).

Back

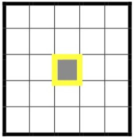
SHI Construction. Calculating SHI for each Cell



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



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

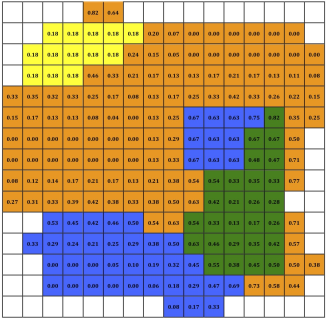


(C) SHI for given cell is $1 - 7/24 = 0.67$

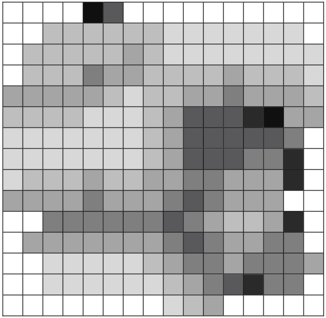
Note: This figure illustrates step (2.) of the county-level SHI construction - calculating the the probability that a randomly selected neighboring cell (in bold-black frame) is of a different soil type than a given cell (in bold-yellow frame).

Back

SHI Construction. SHI for each Cell



(A) Numeric representation

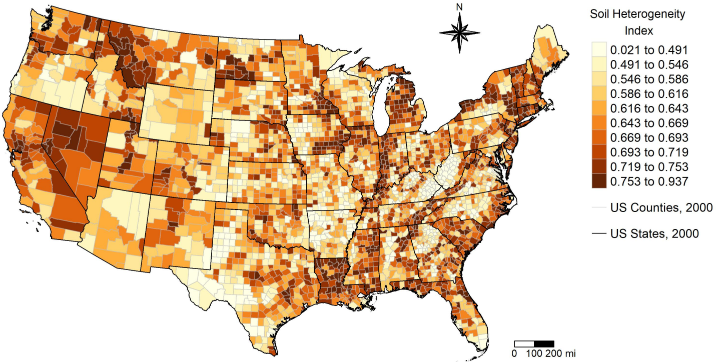


(B) Heat map representation

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.

◂ Back

County-level Soil Heterogeneity Index (SHI)



Note: This figure plots the soil heterogeneity index (SHI) for counties in the contiguous U.S. in 2000. Darker color implies a higher soil heterogeneity.

Back

LNI correlates with communal morality

	Dependent variable:					
	Rel. Importance of Communal Moral Values		Trump Vote Share 2016		Δ [*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. 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) Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al, 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

From soil homogeneity to close-knit communities

- Soil homogeneity $\Longrightarrow$
  can learn from neighbors $\Longrightarrow$
  incentive to invest in relationships with neighbors
  - Contribute to the development of a close-knit community and a high degree of communal interdependence

- Soil homogeneity $\Longrightarrow$
  social learning is effective $\Longrightarrow$
  place greater value on the actions of others

- Soil homogeneity $\Longrightarrow$
  homogeneity in agricultural practices $\Longrightarrow$
  development of tight group norms

Back

From soil homogeneity to close-knit communities

- Soil homogeneity $\Longrightarrow$
  can learn from neighbors $\Longrightarrow$
  incentive to invest in relationships with neighbors
- Soil homogeneity $\Longrightarrow$
  social learning is effective $\Longrightarrow$
  place greater value on the actions of others
  - Contribute to the development of an interdependence psychology
  - A "horizontal" version of the hypothesis in Giuliano and Nunn (2020)
- Soil homogeneity $\Longrightarrow$
  homogeneity in agricultural practices $\Longrightarrow$
  development of tight group norms

Back

From soil homogeneity to close-knit communities

- Soil homogeneity $\Longrightarrow$
  can learn from neighbors $\Longrightarrow$
  incentive to invest in relationships with neighbors
- Soil homogeneity $\Longrightarrow$
  social learning is effective $\Longrightarrow$
  place greater value on the actions of others

"Social Learning Hypothesis"

- Soil homogeneity $\Longrightarrow$
  homogeneity in agricultural practices $\Longrightarrow$
  development of tight group norms

Back

From soil homogeneity to close-knit communities

- Soil homogeneity $\Longrightarrow$
  can learn from neighbors $\Longrightarrow$
  incentive to invest in relationships with neighbors
- Soil homogeneity $\Longrightarrow$
  social learning is effective $\Longrightarrow$
  place greater value on the actions of others

"Social Learning Hypothesis"

- Soil homogeneity $\Longrightarrow$
  homogeneity in agricultural practices $\Longrightarrow$
  development of tight group norms
  - 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

Back

From soil homogeneity to close-knit communities

- Soil homogeneity $\Longrightarrow$
  can learn from neighbors $\Longrightarrow$
  incentive to invest in relationships with neighbors
- Soil homogeneity $\Longrightarrow$
  social learning is effective $\Longrightarrow$
  place greater value on the actions of others

"Social Learning Hypothesis"

- Soil homogeneity $\Longrightarrow$
  homogeneity in agricultural practices $\Longrightarrow$
  development of tight group norms

Back More

Other channels

- Soil heterogeneity $\Longrightarrow$
  greater scope to co-insure
  against adverse agricultural shocks $\Longrightarrow$
  incentive to invest in relationships with neighbors
  - Prediction goes the other direction

- Soil heterogeneity $\Longrightarrow$
  agricultural diversity $\Longrightarrow$
  industrialization, urbanization,
  and innovation (Fiszbein, 2019) $\Longrightarrow$
  loose-knit social structure (Greenfield, 2009)

Back

Other channels

- Soil heterogeneity $\Longrightarrow$
  greater scope to co-insure
  against adverse agricultural shocks $\Longrightarrow$
  incentive to invest in relationships with neighbors

- Soil heterogeneity $\Longrightarrow$
  agricultural diversity $\Longrightarrow$
  industrialization, urbanization,
  and innovation (Fiszbein, 2019) $\Longrightarrow$
  loose-knit social structure (Greenfield, 2009)

Back

Measuring communal ties in historical data

- **A fundamental challenge**: communal ties are unobserved in the data
- **Solution**: proxy close-knit communities with the centrality of communal identity in individuals' self-definition, and infer the later from children naming patterns
  1. The centrality of relationships with in-group members in individuals' self-definition is considered *the fundamental difference* between close- and loose-knit social structures (Hofstede et al., 2010; Triandis, 1995; Markus and Kitayama,1991)
     - Triandis et al. (1990) provide empirical support
  2. Rich literature using naming patterns to identify cultural tendencies in historical data
     - Naming patterns have been used to measure *social identification* with different groups, such as race (Levitt and Fryer, 2004), ethnicity (Fouka, 2019), a nation (Abramitzky et al., 2020), and socioeconomic status (Olivetti and Paserman, 2015)

Back

Local Name Index (LNI): Example, 1940

	Share	LNI	Share	LNI
	Arkansas		Massachusetts	
Billie (boys)	2.03%		0.0005%	

Back

Local Name Index (LNI): Example, 1940

	Share	LNI	Share	LNI
	Arkansas		Massachusetts	
Billie (boys)	2.03%	81.42	0.0005%	0.11

Back

Local Name Index (LNI): Example, 1940

	Share	LNI	Share	LNI
	Arkansas		Massachusetts	
Billie (boys)	2.03%	81.42	0.0005%	0.11
Billie (girls)		77.08		2.43
Billy (boys)		75.74		1.09

Back

Local Name Index (LNI): Example, 1940

	Share	LNI	Share	LNI
	Arkansas		Massachusetts	
Billie (boys)	2.03%	81.42	0.0005%	0.11
Billie (girls)		77.08		2.43
Billy (boys)		75.74		1.09
	Alabama		Pennsylvania	
Waytt (boys)	0.015%	85.45	0.0002%	5.30

Back

Historical association between soil heterogeneity and culture

The estimation framework

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

$$Culture_{ct} = \beta Soil\ Heterogeneity_c + \theta_{s(c)t} + X_c \Gamma + \epsilon_{ct} \qquad (1)$$

  - $Culture_{ct}$ is a cultural outcome of interest in county c in year t
  - $\theta_{s(c)t}$ is a state-by-year fixed effect
  - X_c is a vector of time-invariant geo-climatic controls
  - β is the coefficient of interest

- I cluster at arbitrary grid-cells to account for spatial auto-correlation (Bester et al., 2011)

Back

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 1 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. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

▸ SHI definition ▸ LNI definition ▸ Inference ▸ Religious Diversity ▸ Family Ties

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 1 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. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al, 2011). $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Back

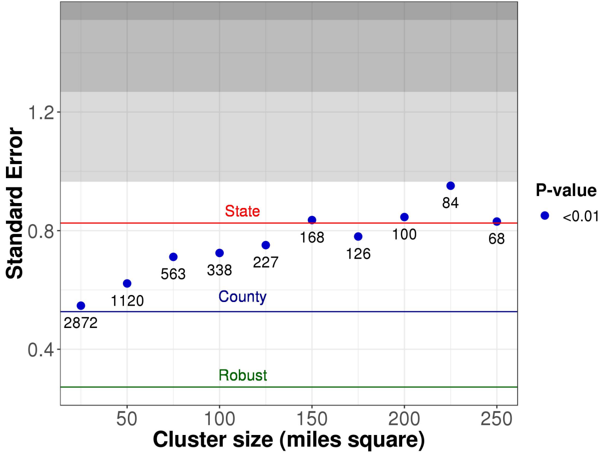
Alternative Definitions of the LNI

	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 1 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. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al, 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

Inference Robustness



Note: This figure plots the standard error of β from the baseline specification of equation 1 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. The data is from the full count censuses between 1850-1940.

Back

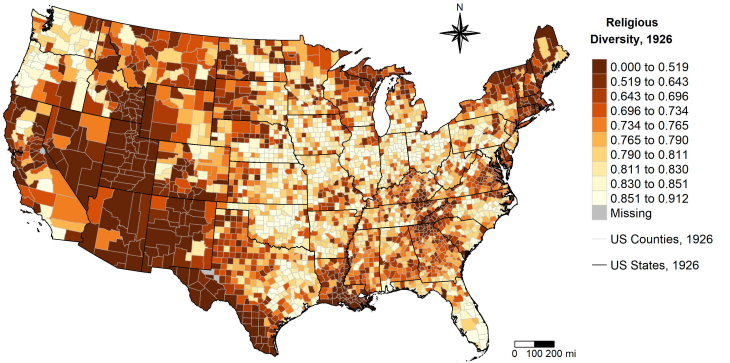
Religious Diversity Index

- I focus on religious diversity as an 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
- I construct a *"Religious Diversity Index"* (RDI) using county-level data on the number of members of religious institutions by denomination between 1850-1926
- The RDI is defined as one minus the Herfindahl–Hirschman Index over the share of members of religious denominations

$$\textit{Religious Diversity Index}_{ct} = 1 - \sum_j s_{cjt}^2$$

Back

Religious Diversity, 1926



Note: This figure plots the county-level *"Religious Diversity Index"* (RDI) in the contiguous U.S. in 1926. Darker color implies higher RDI.

◂ Back

Soil heterogeneity increases the RDI

	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 1 when the dependent variable is the baseline county-level RDI. The data is from the full count census data between 1850-1940. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

The Strength of Family Ties

- I focus on the *"Strength of Family Ties"* (SFT) as an alternative measure for close-knit social networks
- Research in social psychology had identified family ties as a key factor that correlates with interdependence across cultures (Triandis et al., 1990; Vandello and Cohen, 1999; Triandis, 2001)
- In economics, strong family ties correlate with many personal and cultural feature, that in turn were associated with close-knit social networks and interdependence more broadly (Alesina and Giuliano 2010, 2011, 2014; Alesina et al., 2015)

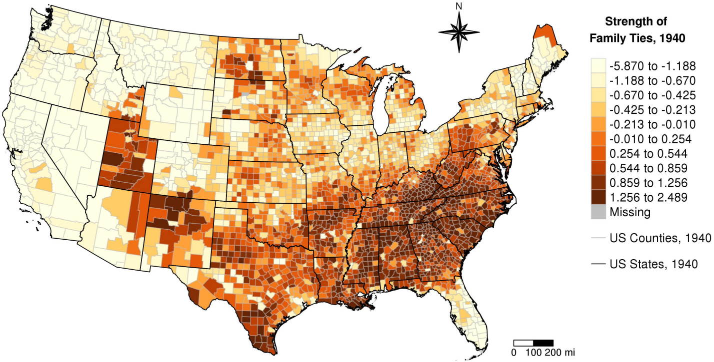
Back

The Strength of Family Ties

- To construct the SFT, I use data on family structure and the choice of living arrangements from the full count census data between 1860-1940
- For each county-year, I calculate
  1. The divorce-to-marriage ratio
  2. The share of elderly people living without a relative
  3. The share of people living with a non-relative
  4. The mean size of families
- Then, for each year, I conduct a PCA using these variables as inputs. The SFT is the first eigenvector
  - It 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. 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)

Back

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.

Back

Soil heterogeneity reduces the SFT

	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 1 when the dependent variable is the baseline county-level SFT. The data is from the full count census data between 1850-1940. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

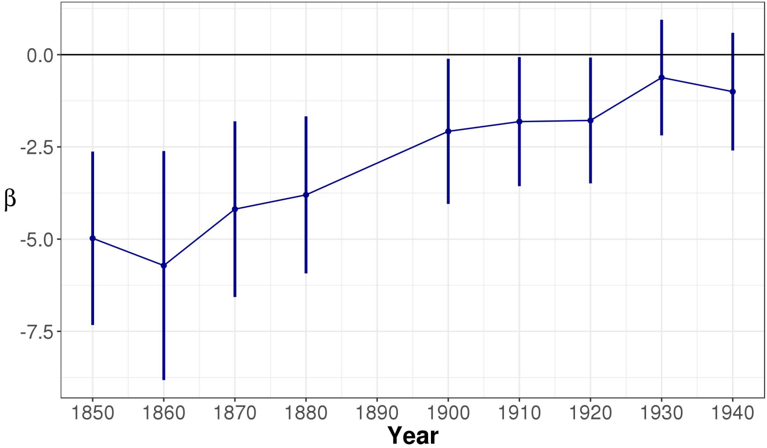
Back

Persistence and long-run impact

- SHI's impact on the LNI is weakening over time ▸ Figure
  - Consistent with soil heterogeneity mattering more when the frontier was settled, and a larger share of the population was engaged in agriculture
- SHI is associated with communal moral values in the long-run ▸ Evidence
  - Consistent with cultural persistence. Soil heterogeneity shaped the nature of social relationships at "critical juncture" in history

◂ Back

SHI's impact on the LNI, period-by-period



Note: This figure reports point estimates and 95% confidence intervals of β from equation 1 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. Regressions also control for state fixed effects and geo-climatice characteristics. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al, 2011).

Back Table

SHI's impact on the LNI, 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 1 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. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al, 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

Long-run impact on communal moral values

I use contemporary survey data to document a persistent effect

- The psychologist Jonathan Haidt (2008) distinguished between two main cultural approaches to morality, and five psychological "foundations" that correspond to them:
  1. **"Individualizing"** (or "universal") approach
     - 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
     - Foundations - Harm / Care, and Fairness / Reciprocity
  2. **"Binding"** (or "communal") approach
     - The group serves as the fundamental source of moral value and individuals are bind together into larger collectives which they are expected to serve
     - Foundations - In-group / Loyalty, Authority / Respect, and Purity / Sanctity
- To measure the degree to which individuals' moral judgment involves the five foundations Graham et al. (2011) developed the *"Moral Foundations Questionnaire"* and surveyed on `www.yourmorals.org` approximately 242,000 Americans between 2008-2018

Back

The estimation framework

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

$$\textit{Moral value}_{ict} = \beta \textit{Soil Heterogeneity}_c + \theta_{s(c)t} + X_c\Gamma + X_i\Lambda + \epsilon_{ict} \qquad (2)$$

  - $\textit{Moral value}_{ict}$ is the moral value 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 county level geo-climactic controls
  - X_i is a vector of pre-determined individual characteristics fixed effects- age, gender and race
  - β is the coefficient of interest

Back

Soil heterogeneity weakens 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 2 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. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

The formation of close-knit communities

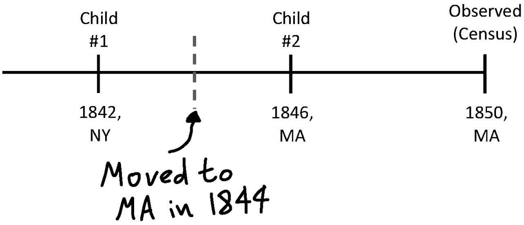
Difference-in-Differences Estimation framework

- Comparing names of children born in families that migrated to counties with varying degrees of soil heterogeneity, before and after the family had moved

The formation of close-knit communities

Difference-in-Differences Estimation framework

- Comparing names of children born in families that migrated to counties with varying degrees of soil heterogeneity, before and after the family had moved



The formation of close-knit communities

Difference-in-Differences Estimation framework

- Comparing names of children born in families that migrated to counties with varying degrees of soil heterogeneity, before and after the family had moved

$$LNI_{ibfc} = \sum_{b=-5^+}^{7^+} \beta_b \cdot \delta_b \cdot Soil\ Heterogeneity_c + \delta_b + \theta_{f(i)} + X_i\Omega + \epsilon_{ibfc} \quad (3)$$

  - LNI_{ibfc} is the LNI score with "local" defined as the state of child i, 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
  - θ_f is a family fixed effect
  - X_i is a potential vector of child i characteristics
  - β_b is the coefficients of interest

◂ Back ▸ canonical-DD ▸ Triple-Difference ▸ canonical-DDD

“canonical” Difference-in-Differences framework

$$LNI_{ibfc} = \beta \cdot \mathbb{1}\{b > 0\} \cdot Soil\ Heterogeneity_c + \delta_b + \theta_{f(i)} + X_i\Omega + \epsilon_{ibfc} \quad (4)$$

- LNI_{ibfc} is the LNI score with “local” defined as the state of child i, 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
- $\mathbb{1}\{b > 0\}$ is a dummy variable for the post movement period
- θ_f is a family fixed effect
- X_i is a potential vector of child i characteristics
- β is the coefficient of interest

Back

Triple-Difference Estimation framework

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

- LNI_{ibfc} is the LNI score with "local" defined as the state of child i, 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
- $\mathbb{1}\{farmer_f\}$ is a dummy variable which equals 1 if the father is a farmer
- θ_f is a family fixed effect
- X_i is a potential vector of child i characteristics
- β_b is the coefficients of interest

Back

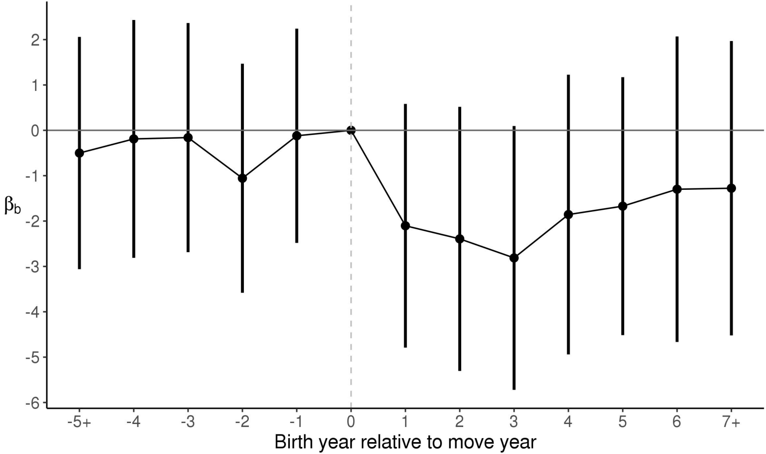
"canonical" Triple-Difference framework

$$
\begin{aligned}
LNI_{ibfc} = & \delta_b + \theta_{f(c)} + \mathbb{1}\{b > 0\} \cdot \mathbb{1}\{farmer_f\} + X_i \Omega + \\
& \gamma \cdot \mathbb{1}\{b > 0\} \cdot Soil\ Heterogeneity_c + \\
& \beta \cdot \mathbb{1}\{b > 0\} \cdot \mathbb{1}\{farmer_f\} \cdot Soil\ Heterogeneity_c + \epsilon_{ibfc}
\end{aligned} \tag{6}
$$

- LNI_{ibfc} is the LNI score with "local" defined as the state of child i, 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
- $\mathbb{1}\{farmer_f\}$ is a dummy variable which equals 1 if the father is a farmer
- θ_f is a family fixed effect
- X_i is a potential vector of child i characteristics
- β_b is the coefficients of interest

Back

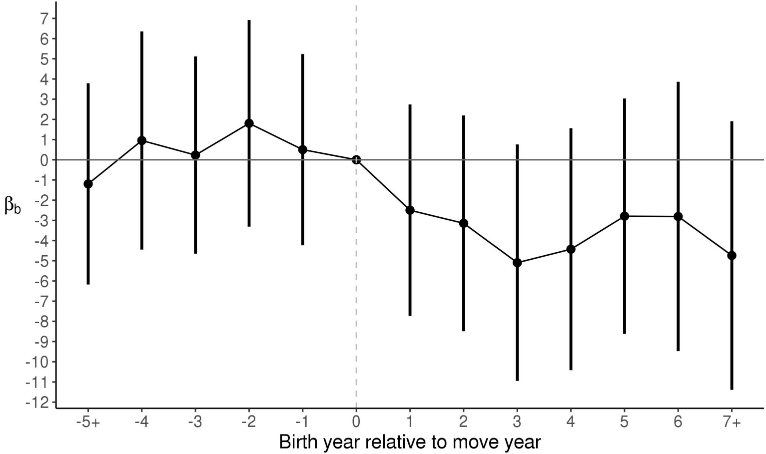
Difference-in-Differences: All Families



Note: This figure plots the estimates of β_b and 95% confidence intervals from equation 3 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. Standard errors clustered at the county in parentheses.

Back

Triple-Difference

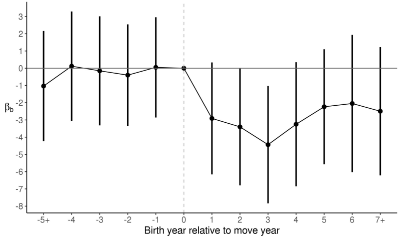


Note: This figure plots the estimates of β_b and 95% confidence intervals from equation 5 when the dependent variable is children's LNI where "local" is defined as the state. The data is from the full count censuses between 1850-1880. Standard errors clustered at the county in parentheses.

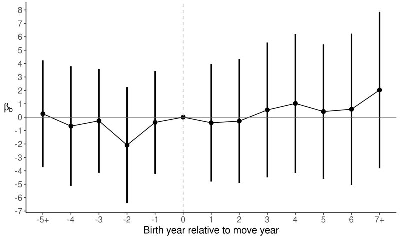
Back

Causal estimates: Difference-in-Differences

Soil heterogeneity only impacts farmers



Farmers' Families



Non-Farmers' Families

Note: This figure plots the estimates of β_b and 95% confidence intervals from equation 3 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. Standard errors clustered at the county in parentheses.

back

Causal estimates: Difference-in-Differences

Soil heterogeneity only impacts farmers

	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 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. Standard errors clustered by county in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

◂ Back ▸ Robustness: sample ▸ Robustness: controls

DD Robustness: Sample

Soil heterogeneity only impacts farmers

	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^{***} (0.800)	-2.802^{***} (0.837)	-0.991 (1.352)
	Panel B: Include all races		
Post Move × Soil Heterogeneity	-2.300^{***} (0.851)	-3.673^{***} (0.931)	0.081 (1.112)
	Panel C: Include multiple moves		
Post Move × Soil Heterogeneity	-1.738^{**} (0.825)	-3.195^{***} (0.875)	0.898 (1.152)
Households Fixed Effects	✓	✓	✓

Note: This table presents estimates of equation 4 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. Standard errors clustered by county in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

DD Robustness: Individual controls

Soil heterogeneity only impacts farmers

	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 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. Standard errors clustered by county in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

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 $\times$ 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 1 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. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

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 1 when the dependent variable is a county-level agricultural diversity index for the years 1880-1935, standardized into z-scores. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

Soil heterogeneity lowers 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 1 when the dependent variable is the county-level growth rate of share of farms using fertilizer, standardized into z-scores. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

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 1 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. Standard errors clustered at arbitrary grid cells of 100 miles square in parentheses (Bester et al., 2011). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Back

Contributions to the literature

1. **The formation of close-knit communities**; the kinship structures hypothesis (Enke et al., 2019; Schulz, 2019; Schulz et al., 2019; Henrich, 2020), the in-group cooperation hypothesis (Talhelm et al., 2014; Buggle, 2018), 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)

   - Provide the first empirical evidence supporting the "*Social Learning Hypothesis*" (Shannon, 1945)

   1.1 The historical roots of culture, cultural persistence and cultural change

   1.2 The individualism-collectivism cleavage

2. Social learning

3. Farmers' settlement of the U.S.

Contributions to the literature

1. The formation of close-knit communities

    1.1 **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)

    1.2 The individualism-collectivism cleavage

2. Social learning

3. Farmers' settlement of the U.S.

Back

Contributions to the literature

1. The formation of close-knit communities

   1.1 The historical roots of culture, cultural persistence and cultural change

   1.2 **The individualism-collectivism cleavage** (Hofstede et al., 2010; Markus and Kitayama, 1991; Triandis, 1995; Gorodnichenko and Roland, 2011, 2017, 2015; Buggle, 2018; Bazzi et al., 2020; Beck-Knudsen, 2019; Enke, 2019, 2020)

2. Social learning

3. Farmers' settlement of the U.S.

Back

Contributions to the literature

1. The formation of close-knit communities

   1.1 The historical roots of culture, cultural persistence and cultural change

   1.2 The individualism-collectivism cleavage

2. **Social learning**; in particular in agriculture (Griliches, 1957; Besley and Case, 1994; Foster and Rosenzweig, 1995; Conley and Udry, 2010) and heterogeneity and social learning (Ellison and Fudenberg, 1993; Munshi, 2004; Yamauchi, 2007)

   - First paper to study the effects of social learning on culture

3. Farmers' settlement of the U.S.

Back

Contributions to the literature

1. The formation of close-knit communities
    1.1 The historical roots of culture, cultural persistence and cultural change
    1.2 The individualism-collectivism cleavage
2. Social learning
3. **Farmers' settlement of the U.S.**; long-run culture and economic development (Bazzi et al., 2020; Fiszbein, 2019; Raz, 2018; Mattheis and Raz, 2019; Smith, 2019).

Back