Geographic axes and the persistence of cultural diversity

David D. Laitin[a,1], Joachim Moortgat[b], and Amanda Lea Robinson[a]

[a]Department of Political Science, Stanford University, Stanford, CA 94305; and [b]Resevoir Engineering Research Institute, Palo Alto, CA 94301

Contributed by David D. Laitin, April 3, 2012 (sent for review December 1, 2011)

Jared Diamond's *Guns, Germs, and Steel* [Diamond J, (1997) *Guns, Germs, and Steel* (WW Norton, NY)] has provided a scientific foundation for answering basic questions, such as why Eurasians colonized the global South and not the other way around, and why there is so much variance in economic development across the globe. Diamond's explanatory variables are: (*i*) the susceptibility of local wild plants to be developed for self-sufficient agriculture; (*ii*) the domesticability of large wild animals for food, transport, and agricultural production; and (*iii*) the relative lengths of the axes of continents with implications for the spread of human populations and technologies. This third "continental axis" thesis is the most difficult of Diamond's several explanatory factors to test, given that the number of continents are too few for statistical analysis. This article provides a test of one observable implication of this thesis, namely that linguistic diversity should be more persistent to the degree that a geographic area is oriented more north-south than east-west. Using both modern states and artificial geographic entities as the units of analysis, the results provide significant confirmation of the relationship between geographic orientation and cultural homogenization. Beyond providing empirical support for one observable implication of the continental axis theory, these results have important implications for understanding the roots of cultural diversity, which is an important determinant of economic growth, public goods provision, local violence, and social trust.

geography | nation-state | culture | languages

In "Spacious Skies and Tilted Axes," a chapter of *Guns, Germs, and Steel*, Jared Diamond begins with a simple observation from which he draws world historical implications (1). Asking his reader to glance at a world map, he assures us that "You'll be struck by an obvious difference. The Americas span a much greater distance north-south (9,000 miles) than east-west; only 3,000 miles at the widest, narrowing to a mere 40 miles at the Isthmus of Panama . . . In contrast, the major axis of Eurasia is east-west." Diamond explores what effect this difference had on human history.

Diamond's biogeographical insight relies on the well-established notion that plant and animal populations have evolved to survive within specific seasonal regimes that, importantly, vary greatly with latitude. Agricultural innovations in domesticating grains and animals can therefore spread more easily in the east-west direction than to the north or south. Because these innovations diffuse in either an easterly or westerly direction, motivated by the population explosions that are facilitated by technological advance, the technologies of production and sophisticated cultural tools, such as writing, follow suit. Although the exact mechanisms of cultural spread remain controversial (2–4), here we will assume that conquest of, or mass migration to, new areas entails the cultural domination of those bringing the innovations, and the disappearance from history of those peoples who are displaced.

This simple observation facilitates an understanding of broad historical patterns. The rapid spread of grains, animals, technologies, and empires across Eurasia is attributed to the low level of variation in latitude along the lines of settlement. Meanwhile, agricultural innovations in the Americas (treated as a single continent) and Africa were slower and less successful. These continents' north-south orientations made technological diffusion inappropriate and counter-productive. Diamond also illustrates the dynamics at the subcontinental level, highlighting the spread of agricultural innovation in south China and across the Sahel as supporting evidence. Continental axes, therefore, are for Diamond one of the four important factors explaining Eurasian world domination. "Around those axes," Diamond concludes, "turned the fortunes of history."

However, the continental axis theory has even broader contemporary relevance. One implication of the theory, as we shall elaborate herein, is that east-west expansion of states delimits cultural diversity but north-south expansion sustains it. Research in the social sciences reveals that cultural diversity is associated with poorer provision of public goods (5, 6), lower levels of interpersonal trust (7), higher probabilities of local violence (8), and lower rates of economic growth (9–12). Indeed, if Diamond's axis theory holds, the identified historical patterns continue to impact society, economy, and polity.

However, Diamond's evidence in support of the axis thesis is only suggestive. Unlike the two other major claims in *Guns, Germs, and Steel*, purporting to account for the variation in economic development across the globe—that is, the susceptibility of local wild plants to be developed for self-sufficient agriculture and the domesticability of large wild animals for food, transport, and agricultural production—the continental axis thesis is not put to rigorous test. First, Diamond's evidentiary base consists of only three continents (Eurasia, Africa, and the Americas) and two noncontinental supplements (the Sahel and south China), making it impossible to assess statistically the implications of continental axes for the spread of human populations and technologies. Second, obvious anomalies are ignored. For example, we are told that potentially efficient diffusions from South to Central America, or from Mexico to the Andes, were halted because of either "hot intervening lowlands" or "intervening tropical climates." However, in a discussion of Bantu expansion (at first west to east, but then inexplicably north to south) we learn that they "leapfrogged over drier areas" (1). Why Bantus but not Amerindians could leapfrog is not investigated.

We seek to provide a more rigorous test of Diamond's intuition by moving beyond his comparison of three continents. To increase the number of observations and allow for a statistical test of the impact of geographic axes, we take the nation-state as the unit of analysis. For each country, we measure the degree to which its axes run across ecological biomes relative to within them. This measure takes the form of the ratio of the average length of the north-south axis relative to the average length of the east-west axis. To the extent that the ratio is greater, we ask whether there has been greater persistence of cultural diversity within the boundaries of that state. Diamond's theory would expect linguistic diversity to erode naturally with east-west migrations. The combining of societies relying on similar technologies would amalgamate those societies into wider speech communities. The theory would then predict the concomitant loss in viability of either the languages of the migrants or those of societies indigenous to the lands that had been populated by those migrants. Meanwhile, the theory would not expect such erosion with north-south migrations, as in this case even small speech communities would survive culturally intact, relying on their special skills related to conditions in climate zones different from where the migrants left.

POLITICAL SCIENCES

Author contributions: D.D.L. designed research; D.D.L. and A.L.R. performed research; J.M. and A.L.R. analyzed data; and D.D.L. and A.L.R. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: dlaitin@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1205338109/-/DCSupplemental.

This study builds on others that have attempted to provide evidence of the continental axis theory. For example, Turchin, Adams, and Hall find that historical empires were significantly more likely to have expanded farther in the east-west direction than in the north-south direction, but that modern states do not exhibit the same inclination (13). Thus, although they provide evidence that the processes described by Diamond are apparent in historical empires, they make no claims about the implications such expansion had for the populations in its wake. Currie and Mace (14) find that political complexity correlates with the area covered by ethnolinguistic groups, and use Diamond's thesis to argue that the east-west orientation of Eurasia, in conjunction with the highly complex Russian state system, facilitated the rapid expansion of the Russian language zone. Similarly, Cashdan (15) argues that the capacity for state expansion is the determining factor for linguistic expansion.

The present study moves beyond these contributions in two important ways. First, our dataset has the largest coverage to date. While Turchin, Adams, and Hall (13) measure geographic axis ratios for only the 29 largest countries in the world, our analysis includes most modern states, allowing us to control for the size of the country rather than excluding smaller countries. Second, our measure of cultural diversity moves beyond what Cashdan (15) calls the "ethnographic present." We capture the persistence of cultural diversity by measuring the percentage of historically indigenous language groups in all countries' histories that are still represented by a self-sustaining language community today. Thus, although our measure of historically spoken languages has a presentism bias (see *Materials and Methods*), the inclusion of recently extinct and moribund languages allows us to better capture linguistic persistence as a historical process. The result is a systematic dataset on linguistic persistence to complement the important narratives provided by classic studies of language death (16, 17).

The observable implication of Diamond's axis thesis that we test in this article is that if the thesis is correct, the length of a country's north-south axis, relative to its east-west axis, should be positively related to the persistence of cultural diversity. To perform this test, we compiled a dataset that includes data on the geographic axis ratio, degree of linguistic persistence, and several control variables for 147 modern countries. We then use a multivariate linear regression to estimate the effect of geographic axes on cultural persistence while controlling for other factors likely to impact the persistence of cultural diversity.

To calculate the axis ratios, we use Mathematica software to plot the geographic coordinates of a state and measure the lengths of both the north-south and the east-west axes of the main land mass of each country. To capture the relative length of the axes, we calculate the axis ratio by dividing the average length of the north-south axis by the average length of the east-west axis. Thus, an axis ratio equal to one corresponds to a roughly square country; axis ratios above one signify countries with greater north-south orientations; and ratios below one represent more east-west oriented countries. Fig. 1 shows the coordinates and resulting ratio for one country, China. Because of the nature of ratios (i.e., ratios below 1 range from 0 to 1, but ratios above one can theoretically range from 1 to infinity), the measure of relative lengths of axes is positively skewed. Thus, we log-transform the measure of axis ratio, resulting in a normal distribution across countries, and all analyses presented use the natural log of axis ratio.

Our measure of linguistic persistence is conceptualized as the proportion of historically spoken languages within a country that have maintained a self-perpetuating language community. We use Ethnologue (18), the most comprehensive list of human languages, both living and defunct, to construct this variable. After coding all languages for indigeneity and viability (see *Materials and Methods*), we calculate linguistic persistence for each country as the proportion of historically spoken languages persisting in a self-sustaining way today.

In our estimation of the impact of geographic ratios on linguistic persistence, we control for other factors likely to impact the persistence of cultural diversity. First, we control for the total number of historical languages—the original language stock—because our measure of persistence will be lower for countries with fewer historic languages, even if the absolute level of language loss were constant across states. For example, the measure of persistence for a country with only two historic languages will be more impacted by the loss of a single language than a country with 10 historical languages, and we need to control for this differential impact. Second, we control for distance from the equator, because linguistic and cultural heterogeneity has consistently been shown to correlate with proximity to the equator (19, 20), even though there is some evidence that the relationship is less clear for language extinction (21). Third, we include a measure of the range in quality of land for agricultural production. Previous work has found that land quality range is positively related to cultural diversity (22), and theoretically should relate to persistence of diversity for the same reason as geographic orientation; variation in land quality should preclude migration and homogenization in the same way that climatic variation induced by a predominately north-south axis should. Fourth, we include the land area and a proxy for degree of mountainousness of a country, with the expectation that both large size and mountainousness will enable cultural persistence. Finally, we control for the age of a country and expect that the time that a country has had to homogenize its population will be negatively related to the persistence of diversity. We provide the resulting data (Datasets S1 and S2) and summary statistics (Table S1) in the Supporting Information.

Results

Consistent with Diamond's continental axis theory, we find that the degree to which a country is more north-south than east-west is positively related to the persistence of cultural diversity. More specifically, our data show a positive bivariate relationship between the geographic axis ratio of a country and the persistence of its linguistic diversity ($r = 0.13$, $P = 0.11$). The relationship remains positive and the estimated effect size is statistically different from zero at the 1% significance level in a multivariate regression that

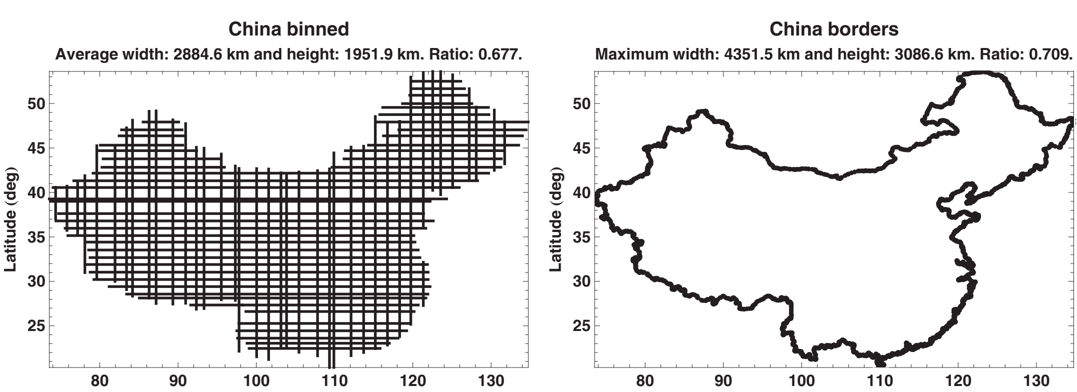

Fig. 1. Visual demonstration of axis ratio calculation for China.

Table 1. Geographic axes and the persistence of linguistic diversity, ordinary least-squares regression

Variable	Model 1	Model 2
Axis ratio (natural log)	0.16**	0.15***
	0.11 (0.04)	0.10 (0.03)
Total no. of languages	0.17***	0.12**
	0.07 (0.02)	0.03 (0.01)
Distance from equator	−0.17**	−0.25***
	−0.32 (0.15)	−0.47 (0.14)
Land quality range	−0.12	−0.18**
	−0.13 (0.10)	−0.18 (0.09)
Area in 1,000s km^2 (natural log)	−0.09	−0.12
	−0.02 (0.01)	−0.02 (0.01)
Elev. diff. in 1,000s m	−0.19**	−0.04
	−0.03 (0.01)	−0.01 (0.01)
Age of country in decades	−0.39***	−0.15*
	−0.02 (0.00)	−0.01 (0.00)
Constant	1.15 (0.06)	1.08 (0.06)
Continent fixed effects?	No	Yes
Observations	147	147
Adjusted r^2	0.40	0.52

Standardized coefficients, ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$; nonstandardized coefficients (robust SEs).

includes control variables (Table 1, model 1). The estimated size of the effect amounts to an increase in the rate of linguistic persistence of roughly 5 percentage points (0.16 SDs) for every one SD increase in the axis ratio. Fig. 2 shows the partial residual plot for the relationship of interest, demonstrating the positive relationship between axis ratio and linguistic persistence, conditional on control variables. The 10 largest outliers in terms of residuals are labeled in Fig. 2, and all countries are labeled in a full-size version of the figure, shown in Fig. S1.

The relationship is robust to adding continent fixed-effects (Table 1, model 2), meaning that the positive relationship between the geographic orientation of a country and the persistence of its cultural diversity holds on average within continents, which themselves have different geographic orientations. Furthermore, the result is robust to excluding any one control variable at a time, although the relationship is only marginally statistically significant without controlling for a country's age, an issue that we discuss

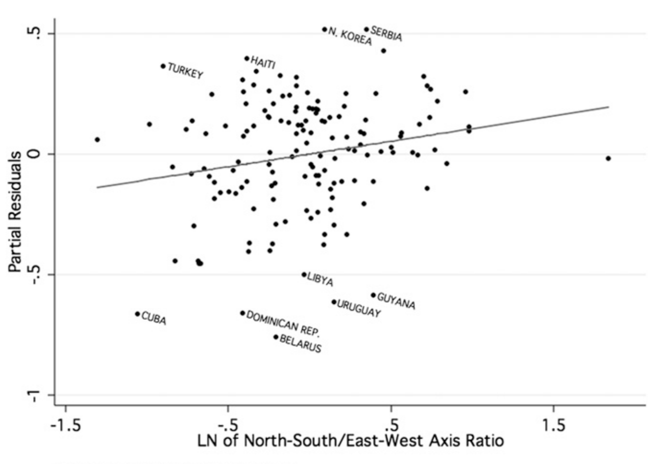

Fig. 2. Partial residuals plot showing the effect of axis ratio on the persistence of linguistic diversity, conditional on control variables (total number of historically spoken languages, distance from the equator, range in land quality, size of the country, elevation difference, and age of the country). Country labels are provided for the 10 largest outliers. A full size version of this figure, with country code labels for all observations, is available in Fig. S1.

below (Table S2). In addition, taking into account the claim of Ricklefs that the climate zones in South America have a more north-south orientation (23), the effect is also robust to calculating axis ratio differently for countries in South America. Ricklef's observation has led past researchers to drop South American cases (13); instead, we replicate our results while taking this anomaly into account. If we calculate the axis ratio of South American states as the average length of the east-west axis divided by the average length of the north-south axis, the strength of the relationship between axis ratio and linguistic persistence is increased ($r = 0.29$, $P < 0.01$) and holds once we include control variables (Table S3).

Although using modern day states as the unit of analysis gives us enough observations for statistical analysis, it also complicates the interpretation of the results. The primary concern with using modern states is that the borders of those states, and thus the relative lengths of their geographic axes, are not exogenous to our outcome of interest. This is because state makers have had agency both in determining the borders of their states and in influencing cultural homogenization in them via nation-building. Like Turchin, Adams, and Hall (13), we find no evidence that modern day states are more likely to run in the east-west direction using a one-sample t test [$\ln(\text{ratio}) = -0.03$, $t(146) = 0.8$, $P = 0.4$]. Thus, the impact of state policy on the persistence of diversity is only problematic for our analyses if state-builders enact differential policies based on geographic axes. If state-builders enact the same policy toward cultural homogenization regardless of the orientation of their state, then the cultural homogenization that occurs because of state agency can be safely ignored in estimating the impact of geographic axes, because nation-building will reduce levels of diversity persistence overall but will do so at rate that is constant across states with different geographic orientations.

However, if state-builders who expanded in the east-west direction instituted different nation-building policies than did state-builders who expanded north or south, then the observed relationship between geographic axis and the persistence of diversity will conflate biogeographical influence with political agency. We could imagine that with east-west expansion state builders would not need to accommodate local populations in return for expertise, and would have incentives to impose their culture on all subjects to reduce the transactions costs of rule (24, 25), but in the north-south direction state builders would be required to accommodate local populations and their cultures in exchange for their geographically distinct expertise. In support of this possibility, Barkey (26) shows that as the Ottoman Empire conquered toward the south, it typically relied on the millet system, allowing for local cultural autonomy; but expansion in the easterly direction involved direct rule and policies of cultural assimilation. McNeill's discussions of the expansion of empires in India and China similarly point to the difficulties of cultural expansion across latitudes, and therefore greater cultural diversity along the north-south axes of those empires (27). In this case, state policy should reinforce geographically determined cultural persistence, resulting in our analyses overestimating the true independent effect of geographic axes. However, if state builders adopt different policies contingent on the geographic orientation of their country, but those policies run counter to the influence of geography, our analyses would underestimate the true size of the effect. For example, state builders may approach nation-building more aggressively in the north-south direction, to counteract the biogeographical impact, with the overall effect of nation-building dampening the influence of geographic orientation.

To determine whether and how endogenous borders are impacting our estimated effect of geographic axes on the persistence of cultural diversity, we interact our axis ratio measure with three indicators of the degree to which borders are endogenous: country age, colonial origin, and border artificiality. The intuition is that the shape of a country can be more readily treated as exogenous when it is younger, when the country was a product of colonialization, and when state borders appear to be artificial. Age of a country is measured in terms of years since entering the international system as an independent state. Postcolonial status is a dummy variable, based on data from the Correlates of War

Table 2. Border endogeneity and the effect of geographic axes on the persistence of linguistic diversity, ordinary least-squares regression

Variable	Model 1	Model 2	Model 3
Axis ratio (natural log)	0.42***	0.01	0.10
	0.28 (0.09)	0.01 (0.06)	0.07 (0.06)
Age of country in decades	−0.51***		
	−0.03 (0.00)		
Ratio (LN) × age	−0.27**		
	−0.02 (0.00)		
Postcolonial state		0.52***	
		0.33 (0.04)	
Ratio (LN) × postcolonial		0.05	
		0.05 (0.09)	
Artificiality of borders			0.20*
			0.06 (0.03)
Ratio (LN) × artificiality			0.17*
			0.14 (0.08)
Constant	0.90 (0.04)	0.50 (0.03)	0.63 (0.03)
Observations	147	147	110
Adjusted r^2	0.30	0.29	0.06

Standardized coefficients, ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$; nonstandardized coefficients (SEs). LN refers to the natural log.

dataset (28), indicating whether a state was decolonized by a foreign power since 1945. Artificiality of a state's borders was determined by Alesina, Easterly, and Matuszeski (29) based on the straightness of borders and the degree to which they partition cultural groups.

The results of estimating models with these interaction terms can be found in Table 2. The coefficient on the interaction between axis ratio and age is negative, suggesting that the impact of geographic axes on linguistic persistence decreases with age. This finding is consistent with the argument that state-building counteracts the impact of biogeography on cultural persistence, and that our estimate of the effect size is an underestimate. The interaction term with postcolonial states is positive (although statistically indistinguishable from zero), which suggests if anything that the impact of geographic axes on linguistic persistence is larger in postcolonial states, the borders of which are presumably less endogenous. This finding is also consistent with the expectation that our estimate based on all countries in the world underestimates the true impact of geographic axes. Finally, the interaction of axis ratio and degree to which borders are arbitrary is positive, which means that the more artificial a country's borders are the larger the effect of axis ratio on the persistence of cultural diversity. Taken together, these three analyses suggest that, if anything, the endogeneity of state borders to the persistence of cultural diversity actually produces an underestimation of the true effect of geographic axes.

As a second strategy to deal with the endogeneity of modern day borders, we replicate our analysis using artificial geographic entities as the unit of analysis. Because our dependent variable is comprised of data that is collected only at the country level, we cannot use purely arbitrary geographic units. As a second best option, we create artificial entities by pairing each country in our sample with each of its contiguous neighbors. For these artificial units, the resulting geographic axes are less endogenously related to linguistic persistence via state-building, allowing a cleaner test of the relationship between geographic axes and the persistence of cultural diversity.

Countries with no land neighbors (islands) are excluded from the dataset. This leaves us with 538 artificial entities resulting from all possible land neighbor combinations of the remaining 136 countries in the sample (Datasets S3 and S4). For each artificial geographic unit, we compute the ratio of average north-south to average east-west axis length using the same procedure as for individual countries. Fig. 3 shows this calculation for one example, the artificial geographic unit created by combining Algeria and Niger. By combining language count data for both countries, we compute the persistence of linguistic diversity in the same way as for the individual countries. To calculate the control variables for each neighbor-pair unit, we simply add the stock of historical languages and total land area and take the averages of distance to equator, elevation difference, and age. Across the 538 artificial units, there is a statistically significant positive relationship between the relative length of geographic axes and the degree to which linguistic differences have persisted ($r = 0.19$, $P < 0.000$), which is robust to including control variables and continent fixed-effects in a multivariate regression (Table 3).

This approach overrepresents countries with many neighbors; for example, the United States only appears twice in the dataset, once paired with Canada and once paired with Mexico, but Russia appears 14 times. If the number of neighbors is correlated with the degree to which axis ratios impact the persistence of diversity, then analyzing the universe of cases will produce a biased estimate of the true effect. To the degree that the number of neighbors is correlated with region of the world, the inclusion of continent fixed effects in the multivariate regression reported above will ameliorate this problem. However, we supplement this analysis of the full population with analyses of 100 random samples drawn from it. For each analysis, we draw a sample of 136 units from the 538 possible country-combinations by randomly selecting only one contiguous neighbor for each of the 136 individual countries. Across these 100 samples, the average effect size is 0.06, with an average SD of 0.03. Because of the estimate being lower than in our single country samples along with much higher SDs, the impact of the north-south to east-west axis ratio is statistically significant at the 1% significance level in only four-tenths of the samples. However, as the histogram portrays in Fig. 4, 99 of the 100 samples

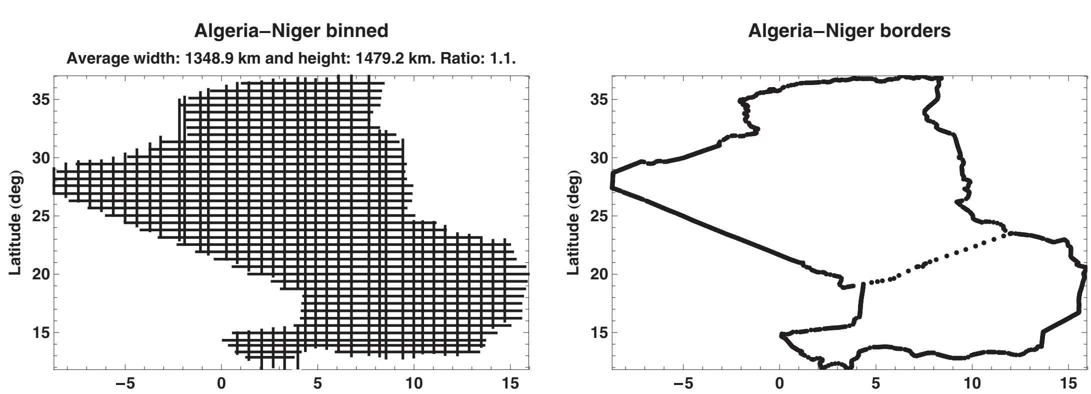

Fig. 3. Visual demonstration of axis ratio calculation for the artificial geographic unit resulting from the combination of Algeria and Niger.

Table 3. Geographic axes and the persistence of linguistic diversity for artificial paired-neighbor units, ordinary least-squares regression

Variable	Model 1
Axis ratio (natural log)	0.08***
	0.05 (0.02)
Total number of indigenous languages	0.14***
	0.03 (0.01)
Average distance from equator	−0.41***
	−0.69 (0.08)
Average land quality range	−0.08**
	−0.11 (0.04)
Area in 1,000s km^2 (natural log)	−0.25***
	−0.05 (0.01)
Average elevation difference in 1,000s m	−0.02
	−0.01 (0.00)
Average age of unit in decades	−0.16***
	−0.01 (0.00)
Constant	1.26 (0.04)
Continent fixed effects?	Yes
Observations	538
Adjusted r^2	0.69

Standardized coefficients, ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$; nonstandardized coefficients (robust SEs).

resulted in a positive estimated coefficient. Combining data from our complete set of paired countries along with the data from 100 randomly paired neighbors suggests endogeneity of modern state borders is not driving our estimate of the biogeographical impact of axis ratio on the persistence of cultural diversity.

Conclusion

We offer here an improved statistical test of Jared Diamond's claim that the continental axes "turned the fortunes of history." By examining an observable implication of the theory at the country-level (instead of Diamond's continent), and computing an axis ratio for all countries, we are able to get sufficient leverage for a statistical test, with far more observations than previous work as well as going beyond tests that examine the diversity of language in a single area. Furthermore, by computing an index of language persistence by country, we are able to go beyond the ethnographic present and capture cultural dynamics.

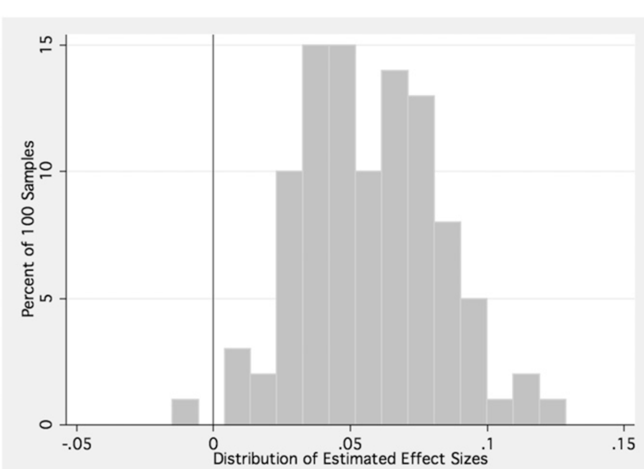

Fig. 4. Histogram of the estimated effect size of axis ratio on linguistic persistence across 100 random samples of artificial paired-neighbor units. Only 1 of the 100 random samples produced an estimated effect size of less than zero (indicated by a horizontal line), and the estimates are centered around 0.05.

Acknowledging the drawback of using modern states, the borders of which are not always exogenous to the processes of cultural homogenization, we supplement this main analysis using two approaches. First, we use indicators of the degree to which state borders are endogenous and evaluate the impact of geographic axes on cultural persistence as a function of that endogeneity. We find suggestive evidence that the more exogenous a state's borders are, the greater the impact of geographic orientation on the persistence of cultural diversity. Next, we evaluate the impact of axis ratio on the persistence of diversity among artificial geographic units comprised of contiguous neighbors, and again find a positive relationship. All of these results provide consistent evidence that the degree to which a geographic unit is oriented more north-south than east-west is related to the persistence of cultural diversity within that unit.

Support for the continental axis theory contributes to our nascent understanding of the roots of contemporary cultural diversity and its social implications. To be sure, we cannot claim from our study that historically conditioned cultural diversity, with the persistence of linguistic heterogeneity, consigns societies to low economic growth (9–12), high rates of generalized distrust of others (7), high likelihood of local violence spiraling beyond original combatants (8), and failure by governments to provide public goods (5, 6). Nonetheless, Diamond's continental axis theory reveals that these unfortunate outcomes are not a reflection of any particular culture's capacities, but rather the historical patterns fostered by geographical constraints that discouraged the integration of cultures.

Materials and Methods

Sample. We start with the 156 countries with populations above one million included in the Fearon and Laitin (30) dataset. We then drop countries that are archipelagos in which the largest land mass does not constitute a majority of the total land mass; we only calculate axis ratios for single land masses to avoid distances covered by water being included in the ratio. This criterion excludes the following nine countries: Fiji, Indonesia, Japan, Malaysia, Mauritius, New Zealand, Papua New Guinea, The Philippines, and Trinidad and Tobago. Ratios for remaining archipelagos, such as Australia, reflect the axis ratio of the main island only.

Axis Ratio. To obtain the north-south vs. east-west axis ratios, we compute the average height divided by the average width of each country and artificial unit. The algorithm is written in Mathematica and uses the CountryData routine sourced by the United States Geological Survey, which provides a list of coordinates defining the country borders. These coordinates are the minimum data required to draw the outlines of each country, but may only have two coordinates defining straight borders, such as the United States–Canada border and the borders of many African countries. In the first step of the computation, additional points are added to borders such that the boundaries are defined by an equal density of points along the borders.

In the second step, the coordinates are binned in both the longitudinal and latitudinal directions into 10–100 bins, depending on the length of the border. The width or height of a bin is computed using the geodesic distance between the minimum and maximum longitude/latitude within each bin, and that difference is converted from degrees to kilometers correcting for the spherical nature of the globe. The average width of a country is derived by averaging the widths of all bins, weighted by the latitudinal extent of each bin. The average height is computed similarly from the binned heights weighted by the longitudinal bin-widths. As a test of the accuracy of this approach, the area of each country is computed twice, by integrating both the widths and the heights of all bins over the whole country. The results are compared with the official country areas in the same data set with very high levels of agreement.

In the third and final step, we calculate the axis ratio by dividing the average north-south axis length by the average east-west axis length.

Cultural Persistence. The dependent variable—persistence of linguistic diversity—is conceptualized as the proportion of historically spoken languages within a country that have maintained a self-perpetuating language community.

To measure the total number of indigenous languages that were historically spoken by a viable speech community within the boundaries of a country, we use data from Ethnologue (18). Languages that meet any of the following criteria were excluded from our count of historically spoken indigenous languages: listed as an immigrant language; originated from a different continent or only spoken on an island other than the mainland; nonspoken (sign) language; used only as a *lingua franca* with no population using it as a first language; or lacking a rural population base. Of the 8,237 languages listed by Ethnologue for the 147 countries in our sample, 6,073 (74%) are indigenous. This list represents a presentism bias, because languages that died out long before Ethnologue began collecting data are not counted. However, language death is a slow phenomenon, and thus many languages that are practically extinct remain coded by Ethnologue as languages spoken only by very few individuals. Thus, even though the editors of Ethnologue warn that they do not include long-extinct languages, we feel that our Ethnologue-based count offers a relatively good proxy for the total number of languages historically spoken within a country.

All languages are then coded as to whether they are presently spoken by a self-perpetuating language community. To code this, we use the UNESCO Atlas of World Languages in Danger (31), which codes languages as "vulnerable," "definitely endangered," "severely endangered," "critically endangered," or "extinct." Languages listed in the Atlas are thus coded as nonself-perpetuating. Languages not listed as in danger by UNESCO are considered to be self-perpetuating, because they are spoken by all generations and there is uninterrupted intergenerational transmission. Of the 6,073 indigenous languages in our sample of 147 countries, 67% have self-perpetuating language communities.

We then divide the total number of viable languages by the total number of indigenous languages historically spoken in that country. Thus, for each of the 147 countries in our dataset, we have a number between 0 and 1 that represents the proportion of historical languages persisting in a sustainable way today.

Control Variables. Our control measure for "original language stock" is the total number of historically spoken languages, described above. For all analyses, this number is divided by 100 to see effect sizes in models. This transformation only affects the coefficients presented, not the size or significance of the effects. "Distance from the equator" is measured in degrees of latitude of the central point in a country, and ranges from 0 (Democratic Republic of Congo) to 64 (Finland). Distance from the equator is also divided by 100 to be able to see the effect size, although again the transformation does not affect the substantive size or statistical significance of the effects. "Variation in land quality" comes from Michalopoulos (22), who first measures the quality of land for agricultural production for small-sized parcels of land. By taking the difference between the highest and lowest quality land parcel within each country, this measure captures the range in land quality. Bahrain exhibits the least variation in land quality (0.002); the United States has the most variation (1.00). Each country's "land area" is measured in square kilometers and logged to normalize the distribution. The "difference in elevation," a proxy for the mountainousness of a country, is the difference in altitude between the highest and lowest points in a country, measured in 1,000 m. A country's "age" is measured by the number of decades since a country entered the international system, which in our dataset begins in 1816. The oldest states in our sample include 13 states at 196 y old, and the youngest include Macedonia and Eritrea at 19 y. Our "postcolonial" measure is an indicator for whether or not a state was decolonized since 1945 (but excluding post-Soviet states that entered the international system since 1991) and comes from the year-entry variable in the Correlates of War dataset (28). In our sample of 147 countries, 63 (43%) are coded as postcolonial. "Artificiality" of a state's borders is coded by Alesina, Easterly, and Matuszeski (29) based on the straightness of borders and the degree to which they partition cultural groups. We use the first principal component of these two variables as our indicator of the degree to which a country's borders are artificial.

ACKNOWLEDGMENTS. Jared Diamond, Peter Turchin, and Stelios Michalopoulos read and commented on earlier versions of this paper; Britta Ellwanger provided research assistance in extracting language count data from the Ethnologue Dataset.

1. Diamond J (1997) *Guns, Germs, and Steel* (WW Norton, New York).
2. Bellwood P (2001) Early agriculturalist population diasporas? Farming, language, and genes. *Annu Rev Anthropol* 30:181–207.
3. Anthony D (2007) *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton Univ Press, Princeton).
4. Renfrew C (1987) *Archaeology and Language* (Jonathan Cape, London).
5. Miguel E (2004) Tribe or nation: Nation building and public goods in Kenya versus Tanzania. *World Polit* 56:327–362.
6. Alesina A, Baqir R, Easterly W (1999) Public goods and ethnic divisions. *Q J Econ* 114: 1243–1284.
7. Putnam R (2007) *E Pluribus Unum*: Diversity and community in the twenty-first century. *Scand Polit Stud* 30:137–174.
8. Fearon J, Laitin D (1996) Explaining interethnic cooperation. *Am Polit Sci Rev* 90:715–735.
9. Nettle D (2000) Linguistic fragmentation and the wealth of nations. *Econ Dev Cult Change* 48:335–348.
10. Easterly W, Levine R (1997) Africa's growth tragedy: Policies and ethnic divisions. *Q J Econ* 112:1203–1250.
11. Pool J (1972) National development and language diversity. *Advances in the Sociology of Language*, ed Fishman J (Mouton, The Hague), pp 213–230.
12. Fishman J (1966) Some contrasts between linguistically homogeneous and linguistically heterogeneous polities. *Sociol Inq* 36:146–158.
13. Turchin P, Adams J, Hall T (2006) East-west orientation of historical empires and modern states. *J World-Syst Res* XII:219–229.
14. Currie TE, Mace R (2009) Political complexity predicts the spread of ethnolinguistic groups. *Proc Natl Acad Sci USA* 106:7339–7344.
15. Cashdan E (2001) Ethnic diversity and its environmental determinants: Effects of climate, pathogens, and habitat diversity. *Am Anthropo New Series* 103:968–991.
16. Dorian NC, ed (1989) *Investigating Obsolescence: Studies in Language Contraction and Death* (Cambridge Univ Press, Cambridge).
17. Crystal D (2000) *Language Death* (Cambridge Univ Press, Cambridge).
18. Lewis MP, ed (2009) *Ethnologue: Languages of the World 16th Edition* (SIL International, Dallas, TX), Available at: www.ethnologue.com. Accessed on March 10, 2011.
19. Nettle D (1999) *Linguistic Diversity* (Oxford Univ Press, Oxford).
20. Collard IF, Foley RA (2002) Latitudinal patterns and environmental determinants of recent human cultural diversity: Do humans follow biogeographical rules? *Evol Ecol Res* 4:371–383.
21. Sutherland WJ (2003) Parallel extinction risk and global distribution of languages and species. *Nature* 423:276–279.
22. Michalopoulos S (2012) Origins of ethnolinguistic diversity. *Am Econ Rev* 102:1508–39.
23. Ricklefs RE (2001) *The Economy of Nature* (WH Freeman, New York).
24. North D (1981) *Structure and Change in Economic History* (Norton, New York).
25. Scott JC (2008) *Seeing Like a State* (Yale Univ Press, New Haven).
26. Barkey K (2008) *Empire of Difference* (Cambridge Univ Press, Cambridge).
27. McNeill WH (1976) *Plagues and Peoples* (Anchor Books, New York).
28. Correlates of War Project (2008) State System Membership List, v2008.1. Available at: http://correlatesofwar.org. Access on January 11, 2011.
29. Alesina A, Easterly W, Matuszeski J (2011) Artificial States. *J Eur Econ Assoc* 9: 246–277.
30. Fearon J, Laitin D (2003) Ethnicity, insurgency, and civil war. *Am Polit Sci Rev* 97:75–90.
31. Moseley D, ed (2010) *Atlas of the World's Languages in Danger.* (UNESCO, Paris) Available at: http://www.unesco.org/culture/en/endangeredlanguages/atlas. Accessed on June 23, 2011.

Supporting Information

Laitin et al. 10.1073/pnas.1205338109

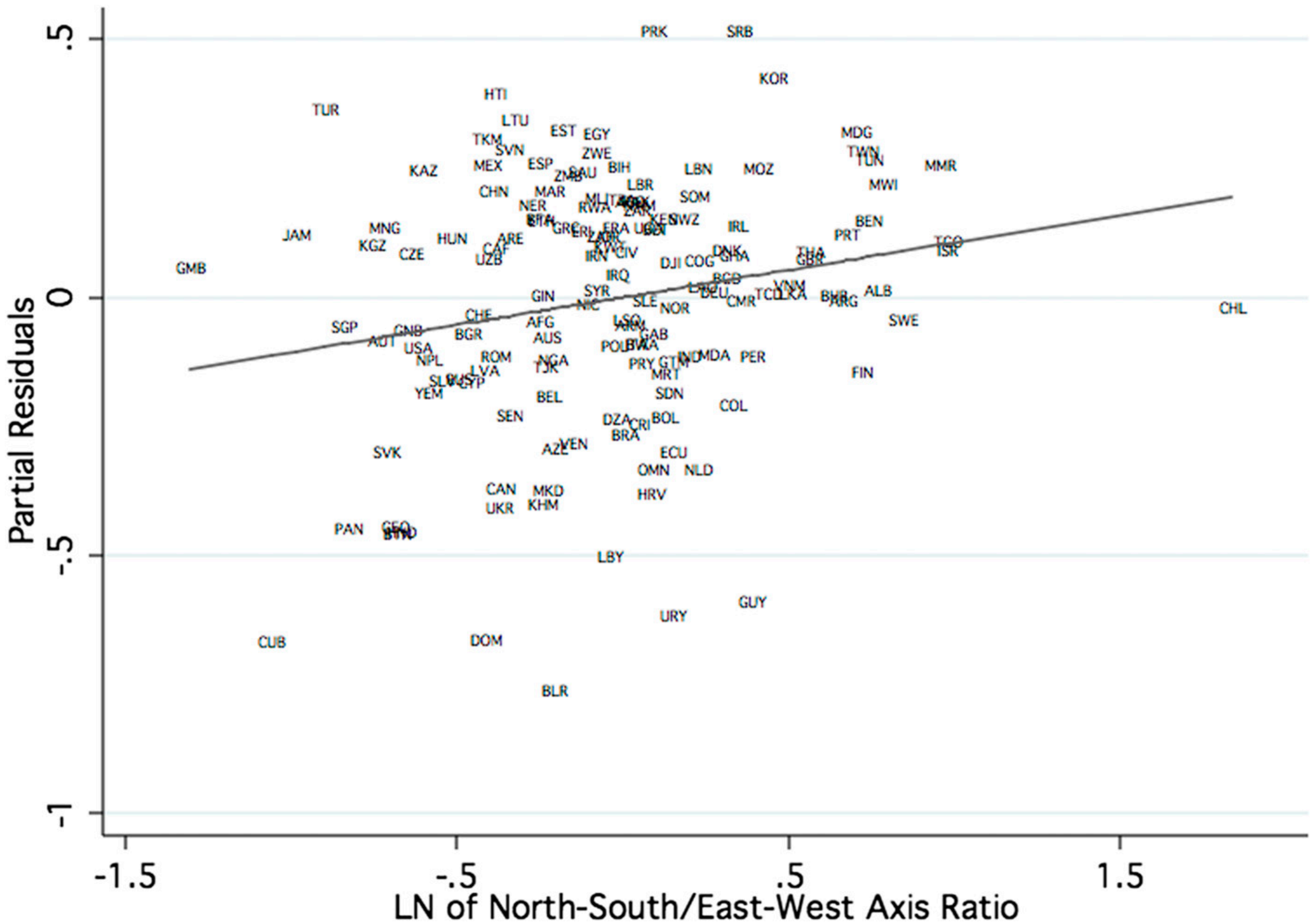

Fig. S1. Partial residuals plot showing the effect of axis ratio on the persistence of linguistic diversity, conditional on control variables (total number of historically spoken languages, distance from the equator, range in land quality, size of the country, elevation difference, and age of the country). Country code labels can be accessed in Dataset S1.

Table S1. Summary statistics

Variable	Mean	SD	Minimum	Maximum	*n*
Axis ratio	1.09	0.65	0.27	6.26	147
Axis ratio, with S. America reversed	1.02	0.49	0.16	2.67	147
Proportion of language persistence	0.64	0.31	0	1	147
No. of self-perpetuating indigenous languages	27.65	59.87	0	493	147
No. of nonself-perpetuating indigenous languages	13.69	29.02	0	169	147
Total no. of indigenous languages	41.34	77.77	0	521	147
Distance from equator (degree latitude)	27.01	16.78	0	64	147
Range in land quality	0.71	0.29	0	1	147
Area (in 1,000s of km^2)	856	2,100	0.49	16,650	147
Elevation difference (1,000 m)	3.07	2.05	0.05	9.00	147
Age, decades since independence	8.83	5.62	1.9	19.6	147
Postcolonial state	0.43	0.50	0	1	147
Artificiality of state borders	−0.01	1.06	−2.10	3.47	110

Table S2. Geographic axes and the persistence of linguistic diversity, sensitivity to excluding control variables one at a time, ordinary least-squares regressions

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Axis ratio (natural log)	0.16**	0.18***	0.15**	0.16**	0.18***	0.12
	0.10 (0.04)	0.12 (0.04)	0.10 (0.04)	0.10 (0.04)	0.12 (0.05)	0.07 (0.05)
Total no. of languages		0.22***	0.16**	0.15***	0.15**	0.18***
		0.09 (0.02)	0.06 (0.02)	0.06 (0.02)	0.06 (0.02)	0.07 (0.02)
Equator distance	−0.22***		−0.20**	−0.17**	−0.16*	−0.27***
	−0.40 (0.14)		−0.37 (0.15)	−0.31 (0.15)	−0.29 (0.15)	−0.50 (0.14)
Land quality range	−0.10	−0.17*		−0.15*	−0.19**	−0.16*
	−0.11 (0.10)	−0.19 (0.10)		−0.16 (0.09)	−0.20 (0.10)	−0.17 (0.09)
Area in 1,000s km^2 (natural log)	−0.03	−0.08	−0.13*		−0.13	−0.15*
	−0.01 (0.01)	−0.01 (0.01)	−0.02 (0.01)		−0.02 (0.01)	−0.03 (0.01)
Elevation difference in 1,000s m	−0.17**	−0.17**	−0.23***	−0.20***		−0.22**
	−0.03 (0.01)	−0.03 (0.01)	−0.03 (0.01)	−0.03 (0.01)		−0.03 (0.01)
Age of country in decades	−0.39***	−0.43***	−0.39***	−0.40***	−0.40***	
	−0.02 (0.00)	−0.02 (0.00)	−0.02 (0.00)	−0.02 (0.00)	−0.02 (0.00)	
Constant	1.12 (0.07)	1.10 (0.06)	1.13 (0.06)	1.11 (0.05)	1.16 (0.06)	1.12 (0.07)
Observations	147	147	147	147	147	147
Adjusted r^2	0.38	0.38	0.39	0.40	0.37	0.27

Standardized coefficients, ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$; nonstandardized coefficients (robust SEs).

Table S3. Geographic axes and the persistence of linguistic diversity, South American axis ratio reversed, ordinary least-squares regression

Variable	Model
Axis ratio (natural log)	0.23***
	0.15 (0.04)
Total number of indigenous languages	0.15**
	0.06 (0.02)
Distance from equator	−0.18**
	−0.34 (0.14)
Land quality range	−0.13
	−0.14 (0.10)
Area in 1,000s km^2 (natural log)	−0.09
	−0.02 (0.01)
Elevation difference in 1,000s m	−0.14*
	−0.02 (0.01)
Age of country in decades	−0.36***
	−0.02 (0.00)
Constant	1.14 (0.07)
Continent fixed effects?	No
Observations	147
Adjusted r^2	0.42

Standardized Coefficients, ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$; nostandardized coefficients (robust SEs).

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)
Dataset S4 (XLSX)