Appendix A. Supplementary Exhibits

Table A1: Summary Statistics

	Mean	Median	Standard Deviation	5th Percentile	95th Percentile	Num. Observations
Panel A: Language Level Variables						
Share of land that is arable	.969	1	.13	.808	1	2,530
Group a estimated utility under trade	2.5	2.32	1.7	.188	5.45	2,530
Group Land Diversity	24,661	10,630	37,997	0	102,007	2,530
Distance between group a, group b centroids.	4.29	4.23	1.11	2.62	6.17	2,530
Distance between language pair families	.621	.667	.278	.114	1	2,530
Group population, 1000s	1,326	14.5	18,649	.15	2,212	2,530
Neighbourhood Area Share (1-100 pct)	11	5.49	14.4	.0614	41.5	2,530
Rank (0-1): Gain From Trade ($\gamma_i = \sum_j^J \frac{\gamma_{ij}}{J}$)	.52	.516	.222	.132	.907	2,530
Rank (0-1): Partner Gain From Trade ($\iota_i = \sum_j^J \frac{\iota_{ij}}{J}$)	.518	.497	.204	.184	.899	2,530
Mean Pairwise Minimum Gains ($\mu_i = \sum_j^J \frac{min\{\gamma_{ij},\iota_{ij}\}}{J}$)	.417	.405	.198	.0828	.777	2,530
Mean Pairwise Interacted Gains ($\mu_i = \sum_j^J \frac{\gamma_{ij} \times \iota_{ij}}{J}$)	.316	.288	.2	.026	.701	2,530
Language Vitality Score	7.1	7	1.72	4	10	2,530
Dominant Language (1/0)	.0391	.	.194	.	.	2,530
Non-dominant Language (1/0)	.633	.	.482	.	.	2,530
Threatened Language (1/0)	.328	.	.469	.	.	2,530
Panel B: Country Level Variables						
Mean Arable Land Share	0.918	0.996	0.168	0.571	1.000	119
Mean Utility Under Trade	2.756	2.560	1.612	0.374	5.334	119
Std.Dev Neighbourhood Area Share	0.069	0.051	0.071	0.000	0.218	119
Std.Dev Neighbourhood Area Share	0.090	0.058	0.100	0.002	0.257	119
Mean Land Diversity	37.146	23.250	43.063	0.447	110.608	119
Mean Utility Gain Interaction	0.313	0.276	0.145	0.135	0.578	119
Mean Utility Gain	0.530	0.509	0.141	0.307	0.785	119
Spatial Inequality	0.506	0.496	0.249	0.110	0.893	119
Ethnic Inequality in Area	0.671	0.717	0.192	0.295	0.895	119
Log Land Area	10.196	10.324	1.644	7.510	12.520	119
Log Population (2000)	16.259	16.155	1.603	13.105	18.804	119

Note: Summary statistics for language and country level variables.

Table A2: Covariate Balance

	Means			Medians		
	Trade Incentives Below Median	Trade Incentives Above Median	p	Trade Incentives Below Median	Trade Incentives Above Median	p
Panel A: Language Level Variables						
Share of group a land that is arable	0.965	0.973	0.130	1.000	1.000	0.790
Group Land Diversity	25,002	24,249	0.612	10,543	10,716	0.928
Panel B: Country Level Variables						
Spatial Inequality	0.475	0.538	0.165	0.485	0.587	0.170
Ethnic Inequality in Population	0.723	0.721	0.960	0.782	0.729	0.333
Ethnic Inequality in Area	0.691	0.650	0.236	0.757	0.660	0.011
Log Land Area	10.179	10.213	0.909	10.223	10.340	0.742
Log Population (2000)	16.448	16.067	0.196	16.215	16.145	0.260
Terrain Ruggedness Index, 100 m.	1.286	1.252	0.877	0.804	0.913	0.857
% Fertile soil	36.451	37.198	0.858	34.143	32.281	0.983
Dummy for landlocked countries	0.317	0.203	0.162	0.000	0.000	0.161
Border artificality measure	31.989	27.974	0.515	20.700	18.000	0.214
Mean Arable Land Share	0.907	0.929	0.479	0.996	0.997	0.669
Mean Utility Under Trade	2.946	2.563	0.197	2.670	2.326	0.331
Abs. Value of Latitude from Equator	28.084	20.019	0.009	30.412	15.365	0.013

Note: Covariate balance for language and country level variables.

Table A3: Fractionalization Measures of Ethnic Heterogeneity

Measure	Acronym	Short Description	Quoted Description
Ethnolinguistic Fractionalization	ELF	Computed as one minus Herfindahl index of population shares of ethnolinguistic groups	*"The ethnolinguistic fractionalization variable (often referred to as ELF) was computed as one minus the Herfindahl index of ethnolinguistic group shares, and reflected the probability that two randomly selected individuals from a population belonged to different groups. We use the same formula, applied to different underlying data, to compute our measures of fractionalization:* $$FRACT_j = 1 - \sum_{i=1}^{N} s_{ij}^2$$ *where s_{ij} is the share of group i ($i = 1, \ldots, N$) in country j."* quoted from Alesina et al. (2003)
Ethnic Fragmentation	EF	This is a version of the fractionalization measure after carefully taking into account the salience of ethnicity and the way the data is constructed and coded.	*"Implicit in the idea of an ethnic group is the idea that members and non-members recognize the distinction and anticipate that significant actions are or could be conditioned on it. So it is natural and perhaps necessary that the "right list" of ethnic groups for a country depend on what people in the country identify as the most socially relevant ethnic groupings. [...] Ideally, the standard for 'the right list' that I am seeking would be defined by a procedure like the following:* 1. *Randomly sample a large number of people in the country.* 2. *Ask each of them to list the major or main ethnic groups in the country.* 3. *Show them or read a list of many possible formulations of the ethnic groups in the country, and ask them to say of which they consider themselves members.* 4. *Repeat (3), asking them to say of which groups on the list most other people in the country would consider them to be members.* 5. *Ask them to try to rank the groups they identified in (3) according to how strongly they identify with the group (e.g., which is 'most important to you,' or some such language).* *[...] Without survey data of this sort, we are forced to review existing lists and secondary sources to apply this standard"* quoted from Fearon (2003, p.198-199)
Cultural Diversity	CD	A version of the fractionalization measure but taking into account the importance of distance between groups as discussed in Desmet et al. (2012).	*"To construct a measure of "cultural fractionalization" analogous to the ethnic fractionalization measure F discussed above, consider drawing two people at random from a country and then computing their expected cultural resemblance, using rij as defined above. In a country with one language group or a set of ethnic groups that all speak highly similar languages, the expected resemblance will be close to 1. In a country with a large number of groups that speak structurally unrelated languages, the expected resemblance will be closer to zero. To get a fractionalization measure analogous to ethnic fractionalization, simply subtract expected cultural resemblance from 1. [...] Formally, cultural fractionalization is $1 - \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j r_{ij}$, where p_i is the proportion of group i and n is the number of groups."* quoted from Fearon (2003, p.212 and p.220)

Note: This table summarizes the main measures of fractionalization used in the literature and gives an intuitive explanation of their construction and the phenomena they are intended to capture.

Table A4: Polarization Measures of Ethnic Heterogeneity

Measure	Acronym	Short Description	Quoted Description
Ethnic Polarization	EP	This measures polarization, which would be maximized with two groups of equal size, unlike fractionalization which is maximized with a large number of small groups.	*"We propose an index of ethnic polarization originally proposed by ReynalQuerol (2002) with the form* $$RQ = 1 - \sum_{i=1}^{N} \left(\frac{1/2 - \pi}{1/2} \right)^2 \pi_i$$ $$RQ = 4 \sum_{i=1}^{N} \pi_i^2 (1 - \pi_i)$$ *The original purpose of this index was to capture how far the distribution of the ethnic groups is from the (1/2, 0, 0, ... 0, 1/2) distribution (bipolar), which represents the highest level of polarization."* quoted from Reynal-Querol and Montalvo (2005) and based on the measure originally proposed in Reynal-Querol (2002). Following Desmet et al. (2012) we define this measure at various levels of aggregation: *This paper seeks to measure linguistic diversity at different levels of aggregation. To do so, we use language trees. We refer to this as a phylogenetic approach (as the linguistics literature does), referring to the fact that tree diagrams capture the genealogy of languages, classified in terms of their family structure.* quoted from Desmet et al. (2012, p.324) We therefore use measures of polarization which account group divisions occurring within the first seven, eleven and fifteen branches of the language family tree.
Peripheral Heterogeneity	PH	This is a type of *social effective antagonism* index that considers the sum over all group-pairs of linguistic distance between the largest (or 'central' group) and peripheral groups (i.e all but the central group), weighted by group sizes.	*"The third [distance] matrix, denoted by* T^c*, assumes there is a center group* c*, such that* $\tau_{jk} = 0$ *if* $j \neq c$ *and* $k \neq c$*. This implies that only the distances between the central group and the other (peripheral) groups matter. [. . .] The A-index is* $$A(0, T^c) = 2 \sum_{j=1}^{K} s_j s_c \tau_{cj}$$ *where the central group* c *is the largest. [. . .] It is important to point out that PH is a variant of GI, with the difference that it takes into account the alienation between the center and the peripheral groups, but not between the peripheral groups themselves."* quoted from Desmet et al. (2009)

Note: This table summarizes the main distribution-based measures of diversity used in the literature and gives an intuitive explanation of their construction and the phenomena they are intended to capture.

Table A5: Correlation between outcomes: Fractionalization and Polarization

(a) Correlation of Fractionalization Measures

	Ethnolinguistic Fractionalization (ELF)	Ethnic Fragmentation (F)	Log. Number of Ethnic Groups	Cultural Diversity (CD)
Ethnolinguistic Fractionalization (ELF)	1.000			
Ethnic Fragmentation (F)	0.743***	1.000		
Log. Number of Ethnic Groups	0.593***	0.460***	1.000	
Cultural Diversity (CD)	0.407***	0.432***	0.676***	1.000

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Ethnolinguistic Fractionalization is from Alesina et al. (2003), Ethnic Fragmentation is from Fearon (2003), Log. number of ethnicities is from Alesina et al. (2016) and Cultural Diversity is based on Fearon (2003) and the original construction in Greenberg (1956).

(b) Correlation of Polarization Measures

	Ethnic Polarization (EP)			Peripheral Heterogeneity Index (PHI)
	Level 7	Level 11	Level 15	
Ethnic Polarization (EP), Level 7	1.000			
Ethnic Polarization (EP), Level 11	0.791***	1.000		
Ethnic Polarization (EP), Level 15	0.788***	1.000***	1.000	
Peripheral Heterogeneity Index (PHI)	0.553***	0.504***	0.505***	1.000

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Ethnic Polarization measured at different levels of aggregation is from Desmet et al. (2012) and Peripheral Heterogeneity Index is from Desmet et al. (2009).

Table A6: Complex Measures of Ethnic Heterogeneity

Measure	Acronym	Short Description	Quoted Description
Ethnic Inequality	EI	Captures inequality across ethnic groups using a Gini coefficient computed using luminosity to measure mean income in an ethnic group's homeland	*"We proxy the level of economic development in ethnic homeland i with mean luminosity per capita, y_i; and we then construct an ethnic Gini coefficient for each country that reflects inequality across ethnolinguistic regions. Specifically, the Gini coefficient for a country's population consisting of n groups with values of luminosity per capita for the historical homeland of group i, y_i, where $i = 1$ to n are indexed in nondecreasing order ($\leq y_{i+1}$), is calculated as follows:* $$G = \frac{1}{n}\left[n + 1 - 2\frac{\sum_{i=1}^{n}(n+1-i)y_i}{\sum_{i=1}^{n} y_i}.\right]$$ *The ethnic Gini index captures differences in mean income—as captured by luminosity per capita at the ethnic homeland—across groups."* quoted from Alesina et al. (2003)
Ethnic Segregation	ES	This measures the degree to which groups are spatially segregated. We use the theoretically correct version which accounts for small subgroups that are counted as 'Other' in the data.	*"... we define our baseline index of segregation for country i as follows:* $$S^i = \frac{1}{M^i - 1}\sum_{m=1}^{M^i}\sum_{j=1}^{J^i}\frac{t_j^i}{T^i}\frac{\left(\pi_{jm}^i - \pi_m^i\right)^2}{\pi_m^i}$$ *where T^i is the total population of country i, t_i^j is the population of region j in country i, and J^i is the total number of regions in country i. [...] In particular, π_m^i is the fraction of group m in country i, and π_{jm}^i is the fraction of group m in region j of country i. [...] One important issue is how to handle the category "other." [...] assume that the group "others" is composed of a number of distinct and small subgroups O that data availability does not permit us to properly classify. Assume also that there is no segregation within the "other" category, i.e., the subgroups of the "other" category are uniformly distributed across all regions. Denote the number of identified groups by n. Then, under these assumptions, one can rewrite the formula for the segregation index S as follows:* $$\hat{S} = \frac{1}{N + O - 1}\left(\sum_{m=1}^{N}\sum_{j=1}^{J}\frac{t_j}{T}\frac{\left(\pi_{jm} - \pi_m\right)^2}{\pi_m} + S_O\right)$$ *where* $$S_O = \sum_{j=1} J\frac{t_j}{T}\frac{\left(\pi_{jO} - \pi_O\right)^2}{\pi_O}$$ *The fraction of "others" in the whole population is represented by π_O, and π_{jO} is the fraction of others in the region j. "* quoted from Alesina and Zhuravskaya (2011, p.1880-1881)

Note: This table summarizes two measures of diversity that are not solely based on population of different groups used in the literature and gives an intuitive explanation of their construction and the phenomena they are intended to capture.

Table A7: Trade Incentives and Language Vitality - Alternate Constructions

(a) Minimum Rank as Mutual Pairwise Incentive

		Status Groupings (1/0)		
	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	1.155	0.022	0.236	-0.258
(μ_i, Mutual Trade Incentives)	(0.429)***	(0.036)	(0.130)*	(0.131)**
Trade is Unlikely: Neighbour Doesn't Gain	-0.683	-0.009	-0.137	0.146
(γ_i, Mean Trade Incentives)	(0.297)**	(0.026)	(0.083)*	(0.084)*
Trade is Unlikely: Only Neighbour Gains	-0.498	0.004	-0.107	0.103
(ι_i, Partner Trade Incentives)	(0.267)*	(0.024)	(0.086)	(0.086)

Note: In this version we use $\tilde{\mu}_i = \frac{1}{J}\sum_{j=1}^{J} min\{\gamma_{ij}, \iota_{ij}\}$ as our measure of mean mutual trade incentives. Definition of outcomes, number of observations and control variables are the same as in table 1.

(b) Maximum Mutual Pairwise Incentive

		Status Groupings (1/0)		
	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	0.807	0.071	0.112	-0.182
($\tilde{\mu}_i$, Max Mutual Trade Incentives)	(0.316)**	(0.033)**	(0.096)	(0.094)*
Trade is Unlikely: Neighbour Doesn't Gain	-0.220	-0.014	-0.072	0.086
(γ_i, Max Trade Incentives)	(0.242)	(0.023)	(0.069)	(0.069)
Trade is Unlikely: Only Neighbour Gains	0.471	0.016	0.058	-0.074
(ι_i, Max Partner Trade Incentives)	(0.245)*	(0.024)	(0.073)	(0.073)

Note: In this version we use $\tilde{\mu}_i = max_{j \in J}\{\gamma_{ij} \times \iota_{ij}\}$ as our measure of mean mutual trade incentives. Definition of outcomes, number of observations and control variables are the same as in table 1.

(c) Maximum Mutual Pairwise Incentive, Mean Unilateral

		Status Groupings (1/0)		
	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	1.794	0.117	0.182	-0.299
($\tilde{\mu}_i$, Max Mutual Trade Incentives)	(0.169)***	(0.021)***	(0.050)***	(0.048)***
Trade is Unlikely: Neighbour Doesn't Gain	-1.028	-0.059	-0.111	0.171
(γ_i, Mean Trade Incentives)	(0.193)***	(0.017)***	(0.057)**	(0.057)***
Trade is Unlikely: Only Neighbour Gains	-0.718	-0.038	-0.068	0.106
(ι_i, Mean Partner Trade Incentives)	(0.194)***	(0.019)**	(0.060)	(0.059)*

Note: In this version we use $\tilde{\mu}_i = max_{j \in J}\{\gamma_{ij} \times \iota_{ij}\}$ as our measure of mean mutual trade incentives, but use the means of the unilateral trade incentives. Definition of outcomes, number of observations and control variables are the same as in table 1.

(d) Maximum Mutual Pairwise Incentive, Mean Unilateral using Minimum Gain

		Status Groupings (1/0)		
	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	1.877	0.121	0.191	-0.311
($\tilde{\mu}_i$, Max Mutual Trade Incentives)	(0.180)***	(0.021)***	(0.054)***	(0.052)***
Trade is Unlikely: Neighbour Doesn't Gain	-1.002	-0.057	-0.109	0.165
(γ_i, Mean Trade Incentives)	(0.193)***	(0.017)***	(0.056)*	(0.056)***
Trade is Unlikely: Only Neighbour Gains	-0.790	-0.042	-0.076	0.118
(ι_i, Mean Partner Trade Incentives)	(0.197)***	(0.020)**	(0.061)	(0.060)*

Note: In this version we use $\tilde{\mu}_i = max_{j \in J}\{min\{\gamma_{ij} \times \iota_{ij}\}\}$ as our measure of mean mutual trade incentives, but use the means of the unilateral trade incentives. Definition of outcomes, number of observations and control variables are the same as in table 1.

Table A8: Trade Incentives and Language Vitality - Region-by-Region Results

(a) Latin America & Caribbean

	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	3.501	0.124	0.094	-0.217
(μ_i, Mutual Trade Incentives)	(1.485)**	(0.109)	(0.419)	(0.408)
Trade is Unlikely: Neighbour Doesn't Gain	-2.314	-0.099	-0.032	0.131
(γ_i, Mean Trade Incentives)	(0.966)**	(0.086)	(0.268)	(0.259)
Trade is Unlikely: Only Neighbour Gains	-1.913	-0.078	-0.131	0.209
(ι_i, Partner Trade Incentives)	(0.932)**	(0.068)	(0.256)	(0.249)
Num. Observations	266	266	266	266
R^2	0.272	0.198	0.213	0.217

(b) Sub-Saharan Africa

	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	2.342	0.025	0.574	-0.599
(μ_i, Mutual Trade Incentives)	(0.988)**	(0.085)	(0.307)*	(0.300)**
Trade is Unlikely: Neighbour Doesn't Gain	-1.040	-0.018	-0.260	0.278
(γ_i, Mean Trade Incentives)	(0.606)*	(0.046)	(0.178)	(0.173)
Trade is Unlikely: Only Neighbour Gains	-0.714	0.029	-0.238	0.210
(ι_i, Partner Trade Incentives)	(0.504)	(0.046)	(0.171)	(0.167)
Num. Observations	885	885	885	885
R^2	0.289	0.281	0.156	0.163

(c) South Asia

	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	5.792	0.635	0.679	-1.314
(μ_i, Mutual Trade Incentives)	(3.124)*	(0.627)	(0.897)	(0.782)*
Trade is Unlikely: Neighbour Doesn't Gain	-3.794	-0.356	-0.647	1.002
(γ_i, Mean Trade Incentives)	(1.467)**	(0.269)	(0.482)	(0.435)**
Trade is Unlikely: Only Neighbour Gains	0.288	0.079	-0.023	-0.055
(ι_i, Partner Trade Incentives)	(1.977)	(0.399)	(0.521)	(0.460)
Num. Observations	176	176	176	176
R^2	0.204	0.178	0.224	0.288

(d) East Asia and Pacific

	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	0.468	-0.016	0.078	-0.062
(μ_i, Mutual Trade Incentives)	(0.611)	(0.021)	(0.223)	(0.225)
Trade is Unlikely: Neighbour Doesn't Gain	-0.457	-0.001	-0.050	0.051
(γ_i, Mean Trade Incentives)	(0.425)	(0.014)	(0.134)	(0.135)
Trade is Unlikely: Only Neighbour Gains	-0.081	0.031	-0.054	0.023
(ι_i, Partner Trade Incentives)	(0.356)	(0.018)*	(0.138)	(0.139)
Num. Observations	939	939	939	939
R^2	0.324	0.300	0.184	0.188

Note: In this table we show results separately for the four regions that have at least one hundred and fifty observations, and make up the largest part of our sample. Definition of outcomes and control variables are the same as in table 1.

Table A9: Language Vitality - Modified Thresholds

	First Robustness Thresholds			Second Robustness Thresholds		
	(1) Dominant	(2) Non-Dominant	(3) Threatened	(4) Dominant	(5) Non-Dominant	(6) Threatened
Trade is Likely: Mutual Benefits	0.058	0.280	-0.338	0.055	0.332	-0.387
(μ_i, Mutual Trade Incentives)	(0.057)	(0.158)*	(0.157)**	(0.041)	(0.119)***	(0.115)***
Trade is Unlikely: Neighbour Doesn't Gain	-0.017	-0.165	0.182	-0.025	-0.159	0.184
(γ_i, Mean Trade Incentives)	(0.038)	(0.094)*	(0.094)*	(0.027)	(0.074)**	(0.071)***
Trade is Unlikely: Only Neighbour Gains	-0.026	-0.108	0.133	-0.011	-0.193	0.204
(ι_i, Partner Trade Incentives)	(0.036)	(0.094)	(0.092)	(0.025)	(0.070)***	(0.067)***
Arable Land Share	✓	✓	✓	✓	✓	✓
Land Diversity	✓	✓	✓	✓	✓	✓
Utility Level under Trade	✓	✓	✓	✓	✓	✓
Area Share Controls	✓	✓	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓	✓	✓
Num. Observations	2530	2530	2530	2530	2530	2530
R^2	0.277	0.221	0.242	0.363	0.242	0.225

Note: The unit of observation is a language-group. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In the first set of robustness classifications, we shift the threshold for being a Dominant language by one category, including Wider Communication, Provincial, and National. In the second set of robustness classifications, we move the threshold for being Threatened down one category, including all classes with vitality less than Shifting.

Table A10: Language Vitality - No 'Vigorous' Class

		Status Groupings (1/0)		
	(1) Vitality Score	(2) Dominant Language	(3) Non-Dominant Language	(4) Threatened Language
Trade is Likely: Mutual Benefits	2.723	0.073	0.500	-0.573
(μ_i, Mutual Trade Incentives)	(0.660)***	(0.057)	(0.178)***	(0.180)***
Trade is Unlikely: Neighbour Doesn't Gain	-1.427	-0.032	-0.267	0.299
(γ_i, Mean Trade Incentives)	(0.418)***	(0.038)	(0.105)**	(0.108)***
Trade is Unlikely: Only Neighbour Gains	-1.279	-0.015	-0.259	0.274
(ι_i, Partner Trade Incentives)	(0.377)***	(0.035)	(0.105)**	(0.104)***
Arable Land Share	✓	✓	✓	✓
Land Diversity	✓	✓	✓	✓
Utility Level under Trade	✓	✓	✓	✓
Area Share Controls	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓
Num. Observations	1879	1879	1879	1879
R^2	0.384	0.394	0.284	0.299

Note: The unit of observation is a language-group. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In this table we drop all language groups assigned to the 'Vigorous' language vitality class, which was assigned as the default classification in some cases, and which therefore may have the least accurate information.

Table A11: Country-Mean Trade Incentives and Fractionalization

	Ethnolinguistic Fractionalization (1)	Ethnic Fragmentation (2)	Log Num. Ethnic Groups (3)	Cultural Diversity (4)
Trade is Likely: Mutual Benefits	2.756	2.839	11.942	1.105
($\bar{\mu}_c$, Mutual Trade Incentives)	(1.060)**	(1.024)***	(2.835)***	(0.597)*
Trade is Unlikely: Neighbour Doesn't Gain	-1.108	-1.175	-5.693	-0.682
($\bar{\gamma}_c$, Mean Trade Incentives)	(0.481)**	(0.494)**	(1.676)***	(0.305)**
Trade is Unlikely: Only Neighbour Gains	-1.604	-1.257	-6.296	-0.566
($\bar{\iota}_c$, Partner Trade Incentives)	(0.724)**	(0.617)**	(1.876)***	(0.385)
Ethnic Inequality in Area	✓	✓	✓	✓
Log Area	✓	✓	✓	✓
Log Population (in 2000)	✓	✓	✓	✓
Mean Group Arable Share	✓	✓	✓	✓
Mean Group Trade Utility	✓	✓	✓	✓
Mean Group Land Diversity	✓	✓	✓	✓
Area Share Controls	✓	✓	✓	✓
Num. Observations	119	112	119	119
R^2	0.289	0.283	0.604	0.410

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a country. The variables $\bar{\mu_c}$ (Mean Mutual Trade Incentives), $\bar{\gamma_c}$ (Mean Trade Incentives), and $\bar{\iota_c}$ (Mean Partner Trade Incentives) are constructed as in equation 5. The outcomes variables are described in detail in section 2.B and all represent fractionalization-style measures of whether a country's population is split into many small groups.

Table A12: Trade Incentives and Fractionalization - Additional Controls

	Ethnolinguistic Fractionalization					Ethnic Fragmentation					Log. Num Ethnic Groups				Cultural Diversity				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Trade is Likely: Mutual Benefits	2.504	3.095	3.483	3.264	2.323	2.711	3.280	1.418	2.070	1.528	10.932	10.471	6.869	5.725	0.982	0.929	0.705	0.741	0.148
($\bar{\mu}_c$, Mutual Trade Incentives)	(1.038)**	(0.904)***	(0.794)***	(0.848)***	(0.776)***	(0.979)***	(0.887)***	(0.988)†	(0.831)**	(0.809)*	(2.897)***	(2.794)***	(2.780)**	(2.770)**	(0.609)†	(0.613)†	(0.684)	(0.694)	(0.638)
Trade is Unlikely: Neighbour Doesn't Gain	-0.823	-1.006	-1.412	-1.317	-0.891	-0.895	-1.066	-0.408	-0.691	-0.454	-4.547	-4.404	-3.088	-2.595	-0.542	-0.525	-0.483	-0.499	-0.230
($\bar{\gamma}_c$, Mean Trade Incentives)	(0.455)*	(0.442)**	(0.446)***	(0.469)***	(0.382)**	(0.489)*	(0.457)**	(0.474)	(0.373)*	(0.356)	(1.681)***	(1.651)***	(1.413)**	(1.474)*	(0.307)*	(0.306)*	(0.323)	(0.321)	(0.277)
Trade is Unlikely: Only Neighbour Gains	-1.576	-1.939	-1.955	-1.766	-1.259	-1.370	-1.718	-0.625	-0.923	-0.626	-6.184	-5.901	-4.068	-3.086	-0.553	-0.520	-0.359	-0.390	-0.070
($\bar{\iota}_c$, Partner Trade Incentives)	(0.721)**	(0.719)***	(0.574)***	(0.613)***	(0.473)***	(0.622)**	(0.590)***	(0.607)	(0.528)*	(0.478)	(1.797)***	(1.743)***	(1.816)**	(1.867)	(0.387)	(0.387)	(0.428)	(0.437)	(0.384)
Ethnic Inequality in Area	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Log Area	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Log Population (in 2000)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean Group Arable Share	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean Group Trade Utility	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean Group Land Diversity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Area Share Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Abs. Value of Latitude		✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓
Ethnic Inequality in Population			✓	✓	✓			✓	✓	✓			✓	✓			✓	✓	✓
Spatial Inequality				✓	✓				✓	✓				✓				✓	✓
Log Num. Ethnic Groups					✓					✓									✓
Num. Observations	119	119	119	119	119	112	112	112	112	112	119	119	119	119	119	119	119	119	119
R^2	0.342	0.389	0.591	0.603	0.740	0.352	0.397	0.456	0.599	0.685	0.645	0.646	0.738	0.754	0.434	0.435	0.480	0.481	0.587

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1, † p < 0.15. The unit of observation is a country. The variables $\bar{\mu_c}$ (Mean Mutual Trade Incentives), $\bar{\gamma_c}$ (Mean Trade Incentives), and $\bar{\iota_c}$ (Mean Partner Trade Incentives) are constructed as in equation 5. The outcomes variables are described in detail in section 2.B and all represent fractionalization-style measures of whether a country's population is split into many small groups.

Table A13: Country-Mean Trade Incentives and Polarization

	Ethnic Polarization			Peripheral Heterogeneity
	Level 7	Level 11	Level 15	
	(1)	(2)	(3)	(4)
Trade is Likely: Mutual Benefits	-0.416	0.562	0.514	-0.297
($\bar{\mu}_c$, Mutual Trade Incentives)	(1.058)	(0.937)	(0.933)	(0.502)
Trade is Unlikely: Neighbour Doesn't Gain	0.242	0.126	0.144	0.068
($\bar{\gamma}_c$, Mean Trade Incentives)	(0.462)	(0.446)	(0.446)	(0.240)
Trade is Unlikely: Only Neighbour Gains	-0.494	-1.249	-1.238	-0.025
($\bar{\iota}_c$, Partner Trade Incentives)	(0.667)	(0.630)*	(0.632)*	(0.282)
pvalue H_0: $\hat{\beta}_\mu^{POL} = \hat{\beta}_\mu^{FRAC}$	0.069	0.070	0.062	0.009
Ethnic Inequality in Area	✓	✓	✓	✓
Log Area	✓	✓	✓	✓
Log Population (in 2000)	✓	✓	✓	✓
Mean Group Arable Share	✓	✓	✓	✓
Mean Group Trade Utility	✓	✓	✓	✓
Mean Group Land Diversity	✓	✓	✓	✓
Area Share Controls	✓	✓	✓	✓
Num. Observations	119	119	119	119
R^2	0.183	0.174	0.176	0.222

Note: The unit of observation is a country. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In columns 1-3 the outcome is a measure of Ethnic Polarization from Desmet et al. (2012) at different levels of aggregation, i.e. using different depths or 'levels' of classification in the family tree of languages to aggregate groups. The higher the level, therefore, the more fine-grained classification of groups. Here $\hat{\beta}_\mu^{POL}$ refers to the coefficient on $\bar{\mu}_c$ (Mutual Trade Incentives) from the regression with the standardized z-score of the given measure of polarization as the outcome. Here $\hat{\beta}_\mu^{FRAC}$ refers to the coefficient on $\bar{\mu}_c$ from the equivalent regression with the standardized z-score of fractionalization, computed at the corresponding level of aggregation as the given polarization measure, as the outcome. We compare to the regression with the standardized z-score of the standard ELF measure in Column 1 of table A11 as the outcome in the case of Peripheral Heterogeneity. The pvalues presented for rejecting H_0 show that the impact of mutual trade incentives on polarization is different to to the impact on fractionalization.

Table A14: Language Vitality (Country-Level)

		Share of Languages in Category (0-1)		
	Vitality Score	Dominant Language	Non-Dominant Language	Threatened Language
	(1)	(2)	(3)	(4)
Trade is Likely: Mutual Benefits	-9.165	-3.702	4.489	-0.787
($\bar{\mu}_c$, Mutual Trade Incentives)	(7.357)	(1.242)***	(1.234)***	(1.222)
Trade is Unlikely: Neighbour Doesn't Gain	-0.357	1.133	-2.064	0.931
($\bar{\gamma}_c$, Mean Trade Incentives)	(3.222)	(0.512)**	(0.588)***	(0.580)
Trade is Unlikely: Only Neighbour Gains	11.171	3.020	-3.011	-0.010
($\bar{\iota}_c$, Partner Trade Incentives)	(5.271)**	(0.835)***	(0.802)***	(0.868)
Ethnic Inequality in Area	✓	✓	✓	✓
Log Area	✓	✓	✓	✓
Log Population (in 2000)	✓	✓	✓	✓
Mean Group Arable Share	✓	✓	✓	✓
Mean Group Trade Utility	✓	✓	✓	✓
Mean Group Land Diversity	✓	✓	✓	✓
Area Share Controls	✓	✓	✓	✓
Num. Observations	119	119	119	119
R^2	0.399	0.433	0.316	0.276

Note: The unit of observation is a country. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. In this table the outcomes are country-level aggregates of the language-level vitality measures. The outcome in Column 1 is the average vitality score of language groups in a country. The outcomes in Columns 2-3 are the shares of language groups in that country that fall into each of the Dominant, Non-Dominant and Threatened language vitality categories.

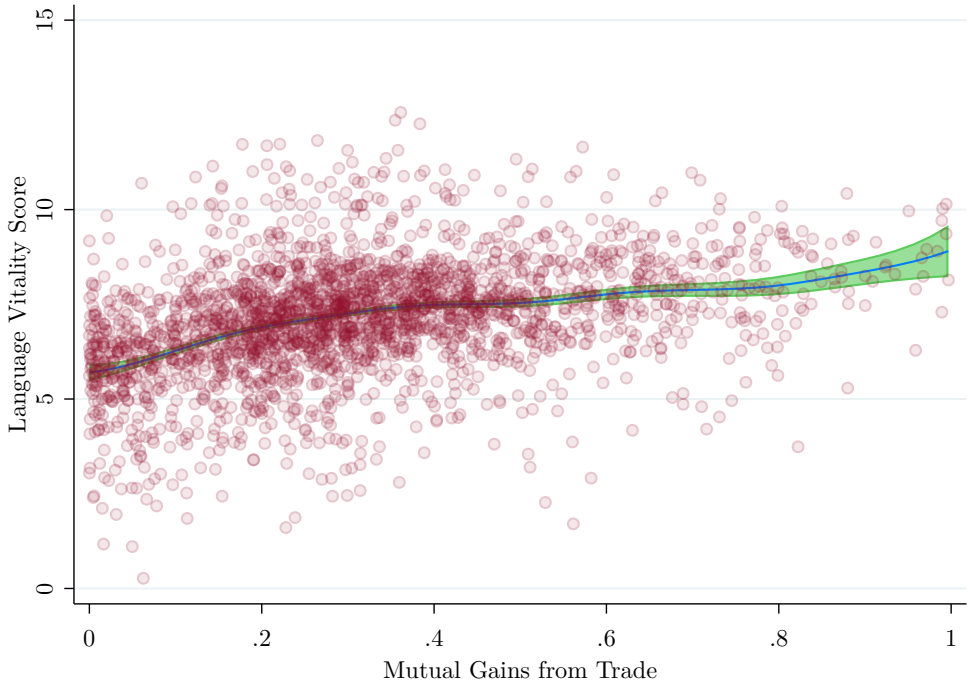
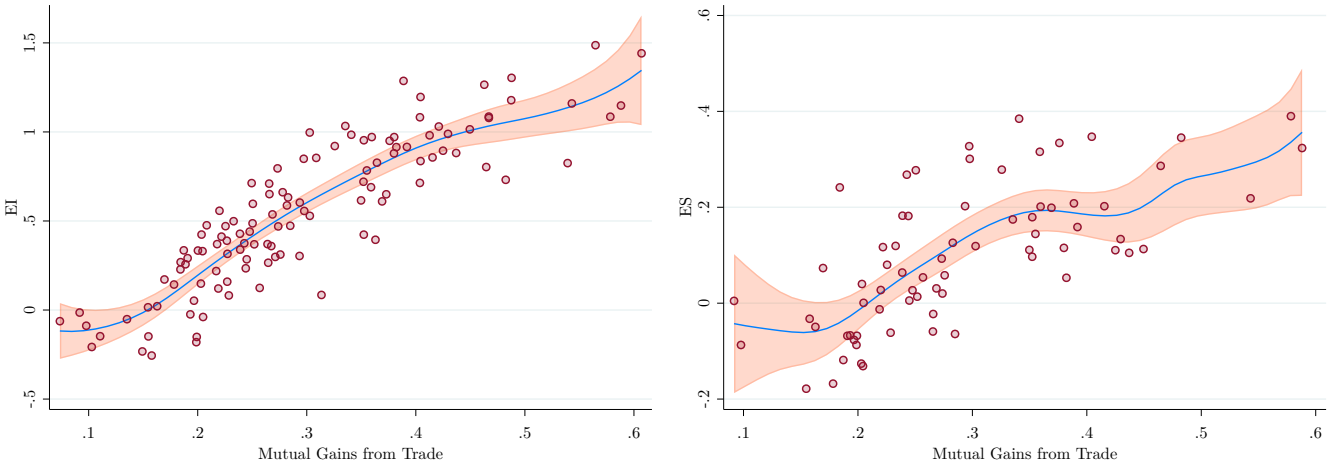


Figure A1: Trade Incentives and Language Vitality

Description: This figure shows the semiparametric relationship between trade incentives and language vitality (where a higher score means higher vitality) at the language level.



(a) Ethnic Inequality (EI) **(b)** Ethnic Segregation (ES)

Figure A2: Trade Incentives and Language Vitality

Description: This figure shows the semiparametric relationship between trade incentives and two measures of ethnic diversity (Ethnic Inequality and Ethnic Segregation). These measures are not simple population identity-based measures, and include other aspects of income and where individuals live, making the conceptual relationship to trade incentives and language survival less clear. Nevertheless, there is a positive and significant relationship between mutual trade incentives and these measures of diversity.

Appendix B. Language Status Data

To measure the status of language groups, we extract scores assigned in the Ethnologue (Lewis, 2009) which follow the Expanded Graded Intergenerational Disruption Scale (EGIDS), developed by Lewis and Simons (2010), which is a more fine-grained version of the original Graded Intergenerational Disruption Scale introduced in Fishman (1991). We describe the coding of the EGIDS in table B1, taken from the Ethnologue website.[17] We scraped this information directly from the Ethnologue website using Python, by accessing the url associated with each three-letter Ethnologue code in our dataset. We then searched for a field named *Language Status* and extracted the associated text. We then checked if the associated string began with one of the categories in the table (*e.g. "6a (Vigorous)"*) and assigned the appropriate variable value if a match was found.[18]

Table B1: Expanded Graded Intergenerational Disruption Scale (EGIDS)

Grouping	Vitality Score	EGIDS Level	Label	Description
Dominant Language	13	0	International	The language is widely used between nations in trade, knowledge exchange, and international policy.
	12	1	National	The language is used in education, work, mass media, and government at the national level.
	11	2	Provincial	The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.
	10	3	Wider Communication	The language is used in work and mass media without official status to transcend language differences across a region.
Non-Dominant Language	9	4	Educational	The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.
	8	5	Developing	The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.
	7	6a	Vigorous	The language is used for face-to-face communication by all generations and the situation is sustainable.
Dominant Language	6	6b	Threatened	The language is used for face-to-face communication within all generations, but it is losing users.
	5	7	Shifting	The child-bearing generation can use the language among themselves, but it is not being transmitted to children.
	4	8a	Moribund	The only remaining active users of the language are members of the grandparent generation and older.
	3	8b	Nearly Extinct	The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language.
	2	9	Dormant	The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency.
	1	10	Extinct	The language is no longer used and no one retains a sense of ethnic identity associated with the language.

Note: This table describes how we map the EGIDS coding of language status in the Ethnologue (Lewis, 2009) into the variable we use in our analysis. The original GIDS scale was developed by Fishman (1991) and expanded into the EGIDS by Lewis and Simons (2010). Descriptions of each category taken from the Ethnologue website: https://www.ethnologue.com/about/language-status

We choose to assign each detailed category a separate variable value as the distinction between them appears to contain relevant information for our analysis. For example, the difference between *6a (Vigorous) "The language is used for face-to-face communication by all generations and the situation is sustainable"* and *6b (Threatened) "The language*

[17] https://www.ethnologue.com/about/language-status

[18] On the Ethnologue website, some coding assessments are marked as a best guess by the Ethnologue editorial team. Exact explanation: *"We use an asterisk as a modifier on the EGIDS estimate to indicate that it represents our editorial best guess. Thus 5* or 6a* indicates a language that we think is most likely to be in vigorous use by all, while 6b* indicates a language that we believe is most likely to be losing speakers."*. We accept these estimates as accurate and so in our data we consider, for example, *6b** and *6b* to be equivalent and assign them the same score.

is used for face-to-face communication within all generations, but it is losing users" is valuable information in terms of language sustainability.[19] We therefore arrive at a 13-point increasing scale for language vitality, with 13 representing the strongest languages of international significance, and 11 representing extinct languages.

After extracting data in this way, we are able to find information on the Ethnologue pages for 6,181 groups. Of these seventeen didn't include a field for *Language Status* or used a non-EGIDS classification and are dropped from the sample.[20]

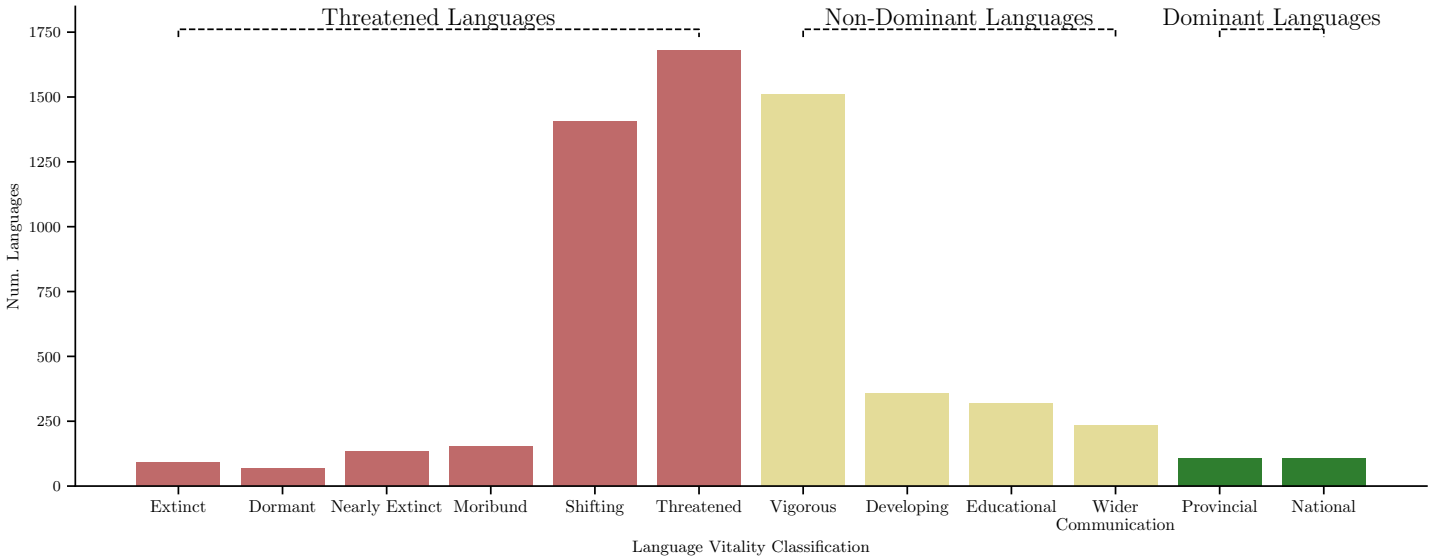


Figure B1: Distribution of Language Vitality Classes

Note: The figure shows the distribution of language vitality classifications.

[19]This importance is also recognized by the editorial board of the Ethnologue: *"From the point of view of sustaining language use, the single most significant break in the EGIDS scale is the divide between 6a and 6b. For languages that are 6a and higher, it is the norm that the language is being learned by all the children within its user community. But at level 6b and below, this is no longer the norm and intergenerational transmission is being disrupted." (quoted from https://www.ethnologue.com/about/language-info*

[20]These non-EGIDS classifications were *9 (Reawakening)* or *9 (Second language only)*.

Appendix C. Overview of Semiparametric Regression Methodology

The semiparametric estimates in this paper use the Verardi and Debarsy (2012) implementation of the Robinson (1988) estimator. In this section we provide a brief overview of the estimator, drawn heavily from Verardi and Debarsy (2012) who provide a more detailed explanation.

The double residual methodology in Robinson (1988) can be used to estimate general models of the following type:

$$y_i = \theta_0 + \mathbf{x}_i\theta + f(z_i) + \varepsilon_i \quad i = 1, \ldots, N \tag{6}$$

where y_i is the dependent variable, $\mathbf{x}_i$ is the vector of variables that enter the model parametrically, and z_i is the variable that enters the model nonparametrically. The first step is to take expectation conditional on z_i of both sides:

$$E(y_i|z_i) = \theta_0 + E(\mathbf{x}_i|z_i)\theta + f(z_i) \quad i = 1, \ldots, N \tag{7}$$

and then subtract this from the original model:

$$y_i - E(y_i|z_i) = [\mathbf{x}_i - E(\mathbf{x}_i|z_i)]\,\theta + \varepsilon_i \quad i = 1, \ldots, N \tag{8}$$

The estimated coefficients $\hat{\theta}$ are then recovered by OLS estimation of the model above after fitting conditional expectations of $\mathbf{x}_i$ conditional on z_i, denoted as $\hat{m}_{\mathbf{x}_i}$:

$$y_i - \hat{m}_y(z_i) = [\mathbf{x}_i - \hat{m}_{\mathbf{x}}(z_i)]\,\theta + \varepsilon_i \quad i = 1, \ldots, N \tag{9}$$

With the estimated coefficients $\hat{\theta}$ in hand, the nonlinear function $f(z_i)$ can be fit by nonparametric estimation of the following model:

$$y_i - \mathbf{x}_i\hat{\theta} = \theta_0 + f(z_i) + \varepsilon_i \quad i = 1, \ldots, N \tag{10}$$

In the semiparametric regressions presented in the figures in this paper, we present exactly these nonparametric fits of $f(z_i)$ where z_i is always the measure of mutual gains from trade (μ_i at the language level and $\bar{\mu}_c$ at the country level).

Appendix D. Synthetic Countries

We supplement our analysis of ethnolinguistic diversity at the country level with a robustness exercise based on the construction of synthetic countries. We show that the relationship between mutual trade incentives and ethnolinguistic fractionalization is robust to using synthetic countries of various sizes. This exercise mitigates the potential concerns raised by the endogenous construction of countries. Some of these concerns include the impact of endogenous size of countries, which has long been associated with economic performance (Easterly & Kraay, 2000; Kuznets, 1960), or the artificiality of borders (Alesina et al., 2011) and partitioning of ethnic groups (Michalopoulos & Papaioannou, 2016). Our approach to artificially constructing cells, and testing sensitivity to a given grid, follows the method outlined in Montalvo and Reynal-Querol (2021).

The first step in our procedure is to divide the area including groups in our sample into a number of cells. We then assign groups to synthetic countries according to which cell their centroid falls into.[21] We then use population figures from the ethnologue to compute a measure of ethnolinguistic fractionalization, following the standard definition of fractionalization,[22] for each of these synthetic countries. As with the fractionalization measure for real countries, this measure is maximized if the synthetic country is made up of a large number of small groups. We generate aggregate control measures from the language-level data in order to replicate our main country-level specification in equation 5 as closely as possible.

We do not take a prior stance on the appropriate size of cell to use, so we begin by dividing the range of latitudes and longitudes into equally-sized intervals. We restrict the range of our group centroids, measured in lat/lon degrees, and divide these ranges into equal intervals. We show, in figure D3 maps of the grids generated by this procedure overlaid on the world map. The most coarse grid comes from dividing the range of latitude/longitude into fourteen intervals, giving $14 \cdot 14 = 196$ cells or potential countries. The most fine grid we use divides the range of latitude/longitude into twenty intervals each for $20 \cdot 20 = 400$ cells or potential countries. Note that only the cells or potential countries that contain at least one group centroid actually end up defining synthetic countries, so the number of synthetic countries created is much lower than the total number of cells.

To ensure the results are robust to where the grid happens to be defined, we again employ a method motivated by Montalvo and Reynal-Querol (2021) and redefine grids by shifting the origin point. We do this by successively shifting the latitudes and longitudes of the grid lines by one quarter of the total interval size.[23] This gives us three alternate

[21] As we have done throughout our analysis, we use the Ethnologue (Lewis, 2009) map to define group homelands, from which we define centroids

[22] See table A3 for additional background on the various measures of fractionalization

[23] Intuitively, this procedure means we are moving the grid 'diagonally' with each variation.

grids of the same size and hence three alternate definitions of synthetic countries defined by grid-cells of the same size.

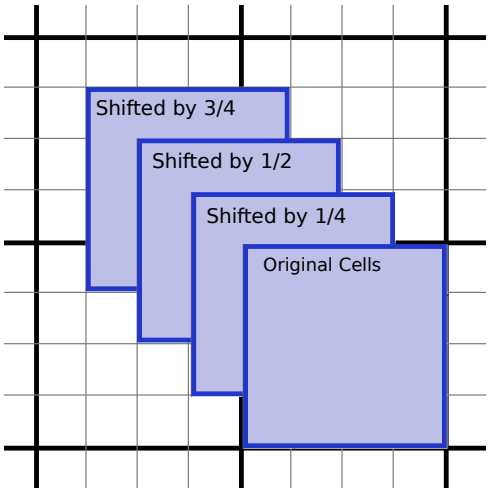
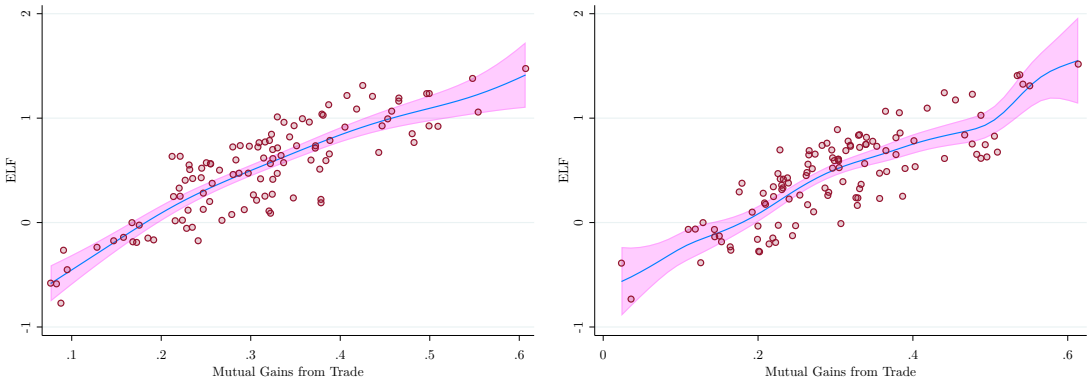


Figure D1: Cells Shifting

Note: This figure gives the intuition for the procedure we use to shift the cells used to define synthetic countries to show robustness to the positioning of cells for a given cell size.

The resolution splitting latitude/longitude ranges into seventeen intervals (resulting in $17 \cdot 17 = 289$ cells) gives us 111 synthetic countries, which is closest to the true number of observations in our cross-country analysis (119). We therefore take this resolution as our main specification, but show robustness to grids that are both larger and smaller.



(a) Original Cells **(b)** Cells Shifted by 1/2 interval

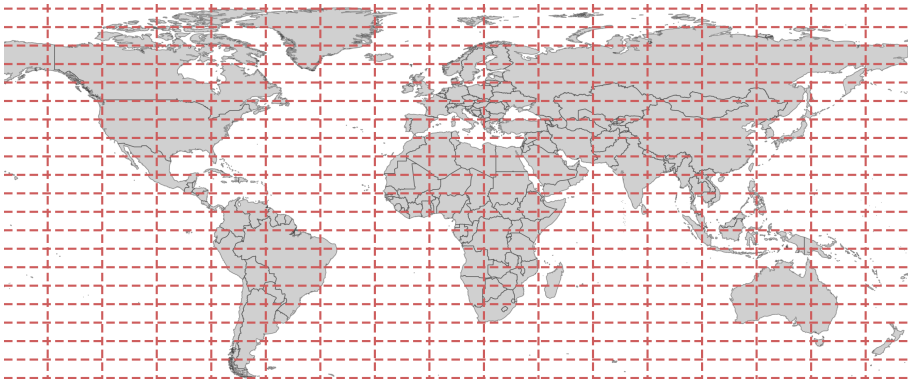
Figure D2: Trade Incentives and Fractionalization in Synthetic Countries

Note: These figures show the semiparametric relationship between trade incentives and Ethnolinguistic Fractionalization (ELF) in synthetic countries.

The positive relationship between mutual trade incentives and national fractionalization holds when we consider these artificially constructed countries. We present the semiparametrically estimated relationship in figure D2 and present the regression estimates for all four variations of the grid in table D1. This relationship is fairly robust to

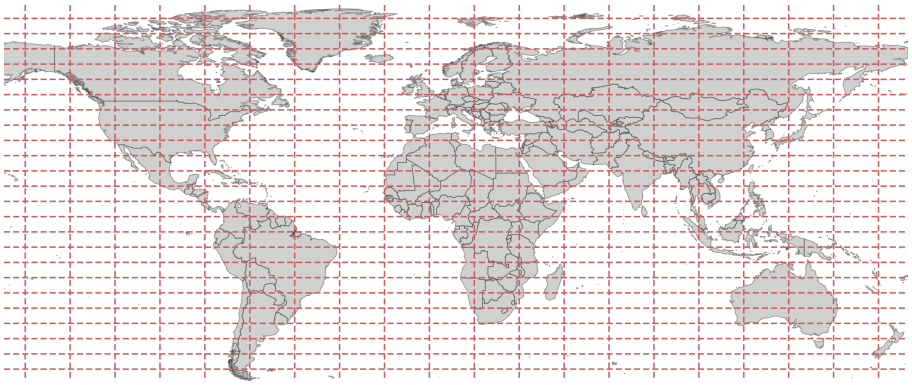
adjusting the size of grid cells to generate larger and smaller numbers of synthetic countries. For four alternative numbers of cells we still find a positive and generally significant relationship (table D2).

These results show that the effect of trade incentives on the vitality of languages significantly impacts fractionalization, even after mitigating concerns related to endogenously sized countries, or colonial borders. This suggests that the impact of threatened languages is important even when we abstract from the impact of national institutions on the vitality and trajectory of language groups.



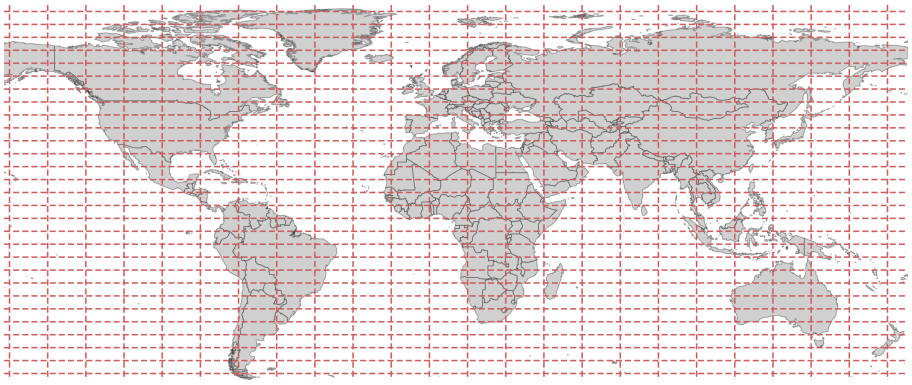
(a) 196 Cells

Note: This maps shows the lines used to split the area of the map containing the group centroids. We divide the area including group centroids into a grid of 14×14 cells for a total of 196 cells.



(b) 289 Cells

Note: This maps shows the lines used to split the area of the map containing the group centroids. We divide the area including group centroids into a grid of 17×17 cells for a total of 289 cells.



(c) 400 Cells

Note: This maps shows the lines used to split the area of the map containing the group centroids. We divide the area including group centroids into a grid of 20×20 cells for a total of 400 cells.

Figure D3: Grid-Cells Defining Synthetic Countries

Table D1: Trade Incentives and Fractionalization with Synthetic Countries

	Original Cells (1)	Cells Shifted by 1/4 (2)	Cells Shifted by 1/2 (3)	Cells Shifted by 3/4 (4)
Trade is Likely: Mutual Benefits	3.773	2.396	1.935	1.876
($\bar{\mu}_c$, Mutual Trade Incentives)	(0.781)***	(0.650)***	(0.936)**	(0.652)***
Trade is Unlikely: Neighbour Doesn't Gain	-1.765	-1.115	-1.197	-1.207
($\bar{\gamma}_c$, Mean Trade Incentives)	(0.501)***	(0.400)***	(0.478)**	(0.419)***
Trade is Unlikely: Only Neighbour Gains	-1.858	-0.984	-0.730	-0.649
($\bar{\iota}_c$, Partner Trade Incentives)	(0.449)***	(0.322)***	(0.440)*	(0.398)
Mean Group Arable Share	✓	✓	✓	✓
Mean Group Trade Utility	✓	✓	✓	✓
Mean Group Land Diversity	✓	✓	✓	✓
Area Share Controls	✓	✓	✓	✓
Num. Observations	111	123	122	111
R^2	0.349	0.254	0.195	0.163

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a synthetic country. In this table our main specification of synthetic countries uses the grid with 289 cells or potential countries, giving the number of synthetic countries that is closest to the number of countries in our sample. In columns 1-4 the synthetic countries are shifted as in figure D1.

Table D2: Synthetic Countries, Robustness

	196 Cells		256 Cells		324 Cells		400 Cells	
	Original (1)	Shifted by 1/2 (2)	Original (3)	Shifted by 1/2 (4)	Original (5)	Shifted by 1/2 (6)	Original (7)	Shifted by 1/2 (8)
Trade is Likely: Mutual Benefits	2.156	2.910	3.277	2.598	1.718	3.149	2.715	1.668
($\bar{\mu}_c$, Mutual Trade Incentives)	(1.364)†	(1.039)***	(0.732)***	(0.798)***	(0.779)**	(0.649)***	(0.732)***	(0.670)**
Trade is Unlikely: Neighbour Doesn't Gain	-1.754	-1.301	-1.997	-1.473	-0.827	-1.774	-1.427	-0.713
($\bar{\gamma}_c$, Mean Trade Incentives)	(0.717)**	(0.617)**	(0.542)***	(0.477)***	(0.468)*	(0.458)***	(0.410)***	(0.473)
Trade is Unlikely: Only Neighbour Gains	-0.113	-2.003	-1.131	-0.756	-0.484	-1.034	-0.919	-1.013
($\bar{\iota}_c$, Partner Trade Incentives)	(0.684)	(0.515)***	(0.313)***	(0.453)*	(0.375)	(0.428)**	(0.393)**	(0.345)***
Mean Group Arable Share	✓	✓	✓	✓	✓	✓	✓	✓
Mean Group Trade Utility	✓	✓	✓	✓	✓	✓	✓	✓
Mean Group Land Diversity	✓	✓	✓	✓	✓	✓	✓	✓
Area Share Controls	✓	✓	✓	✓	✓	✓	✓	✓
Num. Observations	91	96	109	105	124	125	142	137
R^2	0.188	0.336	0.309	0.444	0.172	0.309	0.248	0.169

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is a synthetic country. In this table the size of the grid-cells used to define synthetic countries is different from our main specification, with columns 1 and 2 being more coarse (fewer synthetic countries) and columns 3 and four being more fine (more synthetic countries). The definitions of synthetic countries are also shifted as in figure D1.

References

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, *8*, 155–94.

Alesina, A., Matuszeski, J., & Easterly, W. (2011). Artificial States. *Journal of the European Economic Association*, *9*(2), 246–277.

Alesina, A., Michalopoulos, S., & Papaioannou, E. (2016). Ethnic inequality. *Journal of Political Economy*, *124*(2), 428–488.

Alesina, A., & Zhuravskaya, E. (2011). Segregation and the Quality of Government in a Cross Section of Countries. *American Economic Review*, *101*(5), 1872–1911.

Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of Development Economics*, *97*(2), 322–338.

Desmet, K., Weber, S., & Ortuño-Ortín, I. (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association*, *7*(6), 1291–1318.

Easterly, W., & Kraay, A. (2000). Small States, Small Problems? Income, Growth, and Volatility in Small States. *World Development*, *28*(11), 2013–2027.

Fearon, J. D. (2003). Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, *8*(2), 195–222.

Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Multilingual Matters.

Greenberg, J. H. (1956). The measurement of linguistic diversity. *Language*, *32*(1), 109–115.

Kuznets, S. (1960). Economic growth of small nations. *Economic consequences of the size of nations* (pp. 14–32). Springer.

Lewis, P. (2009). *Ethnologue : Languages of the world*. SIL International.

Lewis, P., & Simons, G. (2010). Assessing endangerment: Expanding Fishman's GIDS [Publisher: Editura Academiei Române].

Michalopoulos, S., & Papaioannou, E. (2016). The long-run effects of the scramble for africa. *American Economic Review*, *106*(7), 1802–1848.

Montalvo, J. G., & Reynal-Querol, M. (2021). Ethnic Diversity and Growth: Revisiting the Evidence. *The Review of Economics and Statistics*, *103*(3), 521–532.

Reynal-Querol, M. (2002). Ethnicity, Political Systems, and Civil Wars. *Journal of Conflict Resolution*, *46*(1), 29–54.

Reynal-Querol, M., & Montalvo, J. G. (2005). Ethnic polarization, potential conflict and civil war. *American Economic Review*, *95*(3), 796–816.

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, *56*(4), 931–954.

Verardi, V., & Debarsy, N. (2012). Robinson's Square Root of N Consistent Semiparametric Regression Estimator in Stata. *The Stata Journal*, *12*(4), 726–735.