Appendix

A   Additional Figures and Tables

Table A.1: Summary statistics, ESS

Variable	Mean	Std	Min	Max	N
Feels discriminated	0.12	0.32	0	1	24,221
Female	0.54	0.50	0	1	24,536
Age	43.89	17.86	13	101	24,442
Contiguous	0.50	0.50	0	1	24,509
Common official language	0.23	0.42	0	1	24,509
Former colonial relation	0.33	0.47	0	1	24,509
Distance between capitals	1.81	2.39	0.06	19.15	24,509
Difference log GDP p.c.	0.76	0.72	0	4.39	24,313
Religious similarity	4.26	2.40	0	9.47	24,536
Distance revenge motifs	0.03	0.03	0	0.29	24,319
Distance violence motifs	0.02	0.02	0	0.17	24,319
Distance cheating motifs	0.04	0.05	0	0.26	24,319
Distance revenge (eHRAF)	0.12	0.29	0	1	9,769
Distance patience	0.42	0.32	0.01	1.56	16,220
Distance risk-taking	0.18	0.18	0	1.32	16,220
Distance positive reciprocity	0.28	0.18	0	1.35	16,220
Distance negative reciprocity	0.28	0.25	0	1.03	16,220
Distance altruism	0.33	0.27	0	1.36	16,220
Distance trust	0.21	0.16	0	1.01	16,220

Table A.2: Pairwise correlations, measures of reputation

	Distance revenge motifs	Distance revenge (eHRAF)	Distance negative reciprocity
Distance revenge motifs	1		
Distance revenge (eHRAF)	0.2011 (0.000)	1	
Distance negative reciprocity	-0.0124 (0.7220)	0.1204 (0.0203)	1

Notes: P-values in parentheses.

Table A.3: Summary statistics, civil conflict

Variable	Mean	Std	Min	Max	N
Conflict incidence	0.09	0.28	0	1	3,969
Distance revenge motifs	0.10	0.08	0	0.41	3,969
Distance cheating motifs	0.11	0.09	0	0.50	3,969
Distance violence motifs	0.06	0.06	0	0.30	3,969
Linguistic distance	0.44	0.29	0	1	3,915
Religious distance	0.34	0.31	0	1	3,969
Geographic distance	0.92	0.72	0	4.02	3,969
Distance jurisdictional hierarchy	1.21	1.06	0	4	3,630
Distance settlement complexity	1.23	1.62	0	6	3,713
Distance High Gods	0.19	0.36	0	1	2,986
Distance herding dependence	10.10	11.90	0	55.75	3,969
Distance segmentary lineage	0.54	0.44	0	1	959
Distance agricultural suitability	0.12	0.12	0	0.44	3,794
Distance ruggedness	509.94	621.96	0	3,234.46	3,794
Distance elevation	519.85	585.06	0	2,791.41	3,795
Distance group size	0.39	0.31	0	0.98	3,969

Table A.4: Differences in revenge prevalence and perceptions of discrimination, two-way clustered standard errors

Dep. variable	Feels discriminated			
	(1)	(2)	(3)	(4)
Distance revenge motifs	0.104**	0.0913*	0.0951**	0.0653**
	(0.0473)	(0.0463)	(0.0464)	(0.0294)
Contiguous		-0.0364***	-0.0393***	-0.0248**
		(0.0120)	(0.00982)	(0.0103)
Common official language		0.0573**	0.0592**	0.0355*
		(0.0216)	(0.0217)	(0.0187)
Former colonial relation		0.00414	0.00543	-0.00418
		(0.0259)	(0.0256)	(0.0241)
Distance between capitals		0.120	0.135*	0.0551
		(0.0718)	(0.0675)	(0.0459)
Difference log GDP p.c.			-0.00565	-0.00901
			(0.0299)	(0.0282)
Religious similarity				-0.153***
				(0.0357)
Observations	23896	23896	23693	23693
R-squared	0.0649	0.0666	0.0664	0.0692

Notes: Data from ESS waves 1-8. An observation is a second generation immigrant. *Feel discriminated* is an indicator for respondents who report being part of a discriminated group. *Distance revenge motifs* measures the absolute difference between the respondent's residence country and their parents' country of birth in the proportion of folklore motifs that are revenge-related. All columns control for gender, age and age squared, host and origin country fixed effects and survey year fixed effects. Standardized beta coefficients reported. Standard errors are clustered two-way at the host and origin country level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5: Differences in revenge prevalence and perceptions of discrimination, interaction of host and origin characteristics

Dep. variable	OLS				2SLS				OLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Feel discriminated											
Revenge motifs	-0.0647***	-0.0345*							-0.0480*		0.0088
	(0.0234)	(0.0179)							(0.0256)		(0.0222)
Negative reciprocity			-0.0474***	-0.0366***	-0.451**	-0.331			-0.0274*	-0.0390	-0.0380
			(0.0173)	(0.0138)	(0.225)	(0.207)			(0.0146)	(0.0294)	(0.0294)
Revenge (eHRAF)							-0227	-0.0121**		-0.0239***	-0.0239***
							(0559)	(0605)		(0710)	(0711)
Patience									-0.318**	-0.0410	-0.0474
									(0.160)	(0.127)	(0.125)
Risk-taking									0.443	0.303	0.309
									(0.395)	(0.434)	(0.436)
Positive reciprocity									0.333	-1.041	-1.035
									(0.440)	(0.724)	(0.728)
Altruism									0.247	0.531	0.530
									(0.222)	(0.375)	(0.373)
Trust									0206	-0.108	-0.136
									(0.294)	(0.409)	(0.415)
Violence motifs									0914	-0398	-0594
									(0.0231)	(0.0282)	(0.0294)
Cheating motifs									0.0128	0.0320	0.0348*
									(0.0164)	(0.0198)	(0.0206)
Observations	23896	23693	15971	15951	15869	15849	9572	9446	15849	6026	6026
R-squared	0.0644	0.0689	0.0563	0.0623			0.0647	0.0677	0.0643	0.0587	0.0588
F-statistic					5.42	5.07					

Notes: Data from ESS waves 1-8. An observation is a second generation immigrant. *Feel discriminated* is an indicator for respondents who report being part of a discriminated group. *Revenge motifs*, *Violence motifs* and *Cheating motifs* are interactions between origin and residence country proportions of folklore motifs that are revenge-, violence- and cheating-related, respectively. *Revenge (eHRAF)* is the interaction between host and origin country values of the eHRAF revenge indicator (averaged across all ethnic groups in a country). *Patience*, *Risk-taking*, *Positive reciprocity*, *Negative reciprocity*, *Altruism* and *Trust* are interactions between host and origin economic preferences from GPS (Falk et al., 2018). All regressions include individual controls for gender, age and age squared, origin and host country fixed effects, and survey year fixed effects, Columns 2, 4, 6, 8-11 also include interactions of bilateral economic, geo-cultural and religious controls from column 4 of Table 2. Standardized beta coefficients reported. Standard errors are clustered at the country pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.6: Sources of variation, civil conflict data

	Nr. obs.	Nr. dyads	First year	Last year
Afghanistan	85	3	1946	2020
Angola	184	4	1975	2020
Benin	19	4	1964	2006
Burundi	31	1	1966	2001
Central African Republic	27	3	1960	2020
China	339	5	1950	2020
Congo	5	1	1964	1968
Congo, Democratic Republic of (Zaire)	559	43	1960	2020
Cote D'Ivoire	64	6	1960	2020
Ethiopia	183	5	1946	2019
Gabon	13	2	1960	2005
Georgia	30	1	1991	2020
Iran (Persia)	150	2	1946	2020
Iraq	17	1	2004	2020
Japan	136	3	1953	2020
Kenya	95	7	1963	2020
Liberia	11	2	1981	2020
Mali	32	1	1960	1991
Niger	163	6	1960	2020
Nigeria	207	10	1960	2020
Rwanda	59	2	1962	2020
Sierra Leone	10	2	1965	2015
Togo	46	2	1960	2020
Uganda	171	24	1967	2020
Zimbabwe (Rhodesia)	21	3	1982	2008

Table A.7: Difference in revenge prevalence and conflict incidence, alternative clustering

Dep. variable	Conflict incidence					
	(1)	(2)	(3)	(4)	(5)	(6)
Distance revenge motifs	0.249	0.329*	0.799*	1.117**	1.121**	1.097**
	(0.122)	(0.093)	(0.082)	(0.015)	(0.014)	(0.010)
Linguistic distance		-0.268	-0.298**	-0.241*	-0.238*	-0.276*
		(0.451)	(0.022)	(0.094)	(0.071)	(0.068)
Religious distance		0.990	2.543	3.576**	3.170**	2.220**
		(0.155)	(0.484)	(0.016)	(0.023)	(0.024)
Log geographic distance		0.421	0.587*	0.564	0.560	0.513
		(0.443)	(0.071)	(0.173)	(0.156)	(0.178)
Distance jurisdictional hierarchy			-0.778	-1.140	-1.132	-1.160
			(0.133)	(0.162)	(0.167)	(0.157)
Distance settlement complexity			-0.0337	-0.159	-0.199	-0.304
			(0.720)	(0.405)	(0.338)	(0.363)
Distance High Gods			0.0866	0.658	0.684*	0.677
			(0.634)	(0.133)	(0.098)	(0.1201)
Distance agricultural suitability				-0.139	-0.153	-0.173*
				(0.116)	(0.108)	(0.0801)
Distance ruggedness				0.982	1.010	1.144
				(0.119)	(0.137)	(0.118)
Distance elevation				-0.345	-0.311	-0.279
				(0.352)	(0.387)	(0.340)
Distance group size					0.443*	1.210*
					(0.054)	(0.065)
Observations	3968	3915	2985	2811	2811	2157
R-squared	0.390	0.393	0.366	0.366	0.368	0.633
Country × Year FE						✓

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Distance revenge motifs* measures the absolute difference between the two ethnic groups in the proportion of folklore motifs that are revenge-related. All columns control for group (government and rebels) fixed effects, country and year fixed effects. Standard errors are clustered at the country level using the wild bootstrap (Cameron et al., 2008). P-values in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.8: Difference in revenge prevalence and conflict incidence, interaction of government and rebel group characteristics

Dep. variable	Conflict incidence					
	(1)	(2)	(3)	(4)	(5)	(6)
Revenge motifs	-0.181	-0.353***	-1.472**	-1.675***	-1.566***	-1.449***
	(0.129)	(0.134)	(0.644)	(0.392)	(0.432)	(0.540)
Linguistic distance		-0.281***	-0.242**	-0.248***	-0.307**	-0.442***
		(0.107)	(0.113)	(0.0753)	(0.119)	(0.140)
Religious distance		0.515	2.540**	3.402***	2.811**	1.712
		(0.521)	(1.215)	(0.881)	(1.255)	(1.296)
Log geographic distance		0.506***	0.716***	0.657***	0.645***	0.623***
		(0.120)	(0.206)	(0.147)	(0.150)	(0.185)
Settlement complexity			-0.0169	-0.161	-0.169	-0.0327
			(0.317)	(0.276)	(0.275)	(0.314)
Jurisdictional hierarchy			-0.0722	0.273	0.309	0.461*
			(0.271)	(0.201)	(0.209)	(0.235)
High Gods			-0.0277	0.149	0.130	0.0706
			(0.211)	(0.282)	(0.285)	(0.341)
Agricultural suitability				-0.100	-0.0622	0643
				(0.104)	(0.118)	(0.121)
Ruggedness				-3.305**	-3.299**	-3.706*
				(1.469)	(1.463)	(2.051)
Elevation				2.424	2.426	2.673
				(1.496)	(1.490)	(2.101)
Group size					-0.489	-1.188***
					(0.708)	(0.394)
Observations	3968	3915	2985	2811	2811	2157
R-squared	0.390	0.393	0.365	0.367	0.368	0.634
Country × Year FE						✓

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Revenge motifs* is the interaction between government and rebel group proportions of folklore motifs that are revenge-related. All other reported estimates are interactions between government and rebel group characteristics. All columns control for group (government and rebels) fixed effects, country and year fixed effects. Standard errors are clustered at the group pair level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.9: Falsification tests, interaction of government and rebel group characteristics

Dep. variable	Conflict incidence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Revenge motifs	-1.449***			-1.423**	-1.705***	-523.1***	-133.5***
	(0.540)			(0.668)	(0.636)	(175.5)	(40.31)
Violence motifs		-0.316		-0.0238			
		(0.285)		(0.208)			
Cheating motifs			2.880	1.968			
			(13.66)	(10.15)			
Dependence on herding					0.801		-9.668***
					(0.846)		(3.361)
Segmentary lineage						41.76***	10.80***
						(14.10)	(3.353)
Observations	2157	2157	2157	2157	2157	550	550
R-squared	0.634	0.633	0.633	0.634	0.634	0.647	0.647

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Revenge motifs, Violence motifs* and *Cheating motifs* are interactions between government and rebel group proportions of folklore motifs that are revenge-, violence- and cheating-related, respectively. *Dependence on herding* is the interaction between government and rebel group in dependence on herding, constructed following Becker (2021) and Cao et al. (2021). *Segmentary lineage* is the interaction between a government and a rebel group indicator for the presence of segmentary lineages, from Moscona et al. (2019). All columns control for group (government and rebels) fixed effects, country by year fixed effects and interactions of bilateral controls from column 6 of Table 6. Standard errors are clustered at the group pair level. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table A.10: Differences in revenge prevalence and conflict incidence, alternative specifications of fixed effects

Dep. variable	Conflict incidence							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance revenge motifs	0.457***	0.498***	0.764***	0.768***				
	(0.123)	(0.128)	(0.246)	(0.193)				
Revenge motifs (interaction)					-0.441***	-0.320*	-1.924***	-1.568***
					(0.134)	(0.172)	(0.589)	(0.440)
Observations	2811	2811	2157	2811	2811	2811	2157	2811
R-squared	0.361	0.365	0.636	0.370	0.368	0.369	0.637	0.373
Powerless group FE	✓	✓	✓		✓	✓	✓	
Year FE	✓	✓			✓	✓		
Country FE		✓				✓		
Country × Year FE			✓	✓			✓	✓
Groups × Country FE				✓				✓

Notes: Data from EPR and UCDP. An observation is an ethnic group pair-country-year. *Conflict incidence* is an indicator for conflict between the government and a paired ethnic group out of power in a given year. *Distance revenge motifs* measures the absolute difference between the two ethnic groups in the proportion of folklore motifs that are revenge-related. Specifications in columns 1-4 include all bilateral distance controls from column 6 of Table 6. Specifications in columns 5-8 include the same controls as interactions between characteristics of the government and rebel group. Standard errors are clustered at the country pair level. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

B Data construction

B.1 Measures of punishment-based reputation

B.1.1 Folklore

Our main measure of punishment-based reputation is the frequency of revenge-related motifs in a society's oral tradition. We rely on the catalog of folklore organized and disseminated by Michalopoulos and Xue (2021). The primary material for this catalog is the work of anthropologist Yuri Berezkin, who compiled a collection of 2,564 motifs in the traditions of 958 groups. A motif "reflects a combination of images, episodes, or structural elements found in two or more texts" (Michalopoulos and Xue, 2021). Michalopoulos and Xue (2021) tagged each motif in the catalog as relating to a concept, using the top 10,000 concepts from ConceptNet, a free online semantic network designed to create word embeddings. A motif is tagged as relevant to a concept if it contains a mention of the concept or of one of the top-50 words returned by ConceptNet when the concept is used as a seed term. A motif may be assigned to different concepts. Michalopoulos and Xue (2021) provide a dataset with the frequency (number) of motifs related to different concepts aggregated at the level of the country and at the level of the ethnic group, for 1,245 groups in the *Ethnographic Atlas* (EA). We rely on these frequencies for the construction of our main independent variable.

We proceed as follows: first, we identify a set of concepts related to revenge and retaliation in the folklore dataset of Michalopoulos and Xue (2021). For consistency with Cao et al. (2021), who also study cultures of honor, we select the following terms: *revenge, punish, punishment, penalty, retaliate, retaliation*. We then compute the total number of motifs tagged by each of the selected keywords and divide that by the total number of motifs catalogued for each group, to adjust for the fact that some societies have richer oral traditions than others. Our final measure is the relative frequency of revenge-related motifs in the folklore of a given society (country or ethnic group).

We follow the same approach to compute the relative frequency of motifs related to violence and deception. For violence, we select the same terms as Cao et al. (2021): *violence, violent, battle, fighting, attach, soldier, guard, troop, army, enemy, fighter, invasion, invade, defender*. For deception, we follow Michalopoulos and Xue (2021) and select the terms *cheat, deceive* and *trick*.

B.1.2 Human Relations Area Files (HRAF)

For the analysis at the country-pair level, we use a complementary measure of punishment-based reputation that draws from the scholarly ethnographic record. We rely on *eHRAF World Cultures*, a database of ethnographic collections covering cultures across the world, produced and maintained by *Human Relations Area Files* (HRAF), a non-profit research agency based at Yale University. The online database provides a summary of the ethnographic record on each culture, organized in sections covering different types of basic information (e.g. location, language), domains of social organization (e.g. marriage and family, inheritance) or activity (e.g. subsistence, trade). We crawl through the list of summaries for any mentions of revenge and retaliation in the two sections most closely related to our theoretical framework: *conflict* and *social control.* The former covers relations to external groups, with a focus on the frequency of war and military activity; the latter relates to inter-community relations, including within-group disputes and sanctions for social deviance. We manually inspect a number of summaries and select the following list of terms as most appropriate to capture the importance of second-party punishment: *veng, feud, vendetta, retali, remedies, retrib, blood, repris, honour, kin, lineage.* We tag as pun-rep all cultures containing at least one of these terms in at least one of the two relevant sections of their culture's summary. We are able to construct this indicator for the 304 societies in eHRAF with recorded information on conflict and social control.

The following examples illustrate what is captured by this method. Societies tagged as pun-rep include some of the most well-known groups that use blood revenge, such as the Pashtun of southeastern Afghanistan and western Pakistan and the Nuer of South Sudan. From the description of social control in the Pashtun society:

"Traditionally social control was maintained by a code of behavior and honor called PAKHTUNWALI. It combines the principles of revenge, hospitality to guests, defense of those who have sought protection in one's care, the chastity of married women, and restraint toward those considered weak or helpless (Hindus, women, and boys). PAKHTUNWALI in some cases contradicts and generally takes precedence over Islamic law. It is harsh – the penalty for illicit sexual behavior, for example, is death – and it is enforced by strong social pressure. Violations of law outside of the activities the code encompasses are dealt with by the JIRGA or the government administration."

And for the Nuer:

"Within the Nuer, homicide is common and is usually related to cattle. Nuer say

that more people have died for the sake of a cow than for any other cause. Acts of homicide can be immediately avenged or held as blood-feuds until such time when the two sides finally square even, and the mechanism to deter homicide and revenge has been the imposition of blood wealth, which is payable in cattle. The norm has been 30 cows paid to the family of the slain person. It can therefore, be said that because cattle are a source of turmoil, a threat of one's cattle being taken away in punishment induces prudence in the relations between people."

Cambodians instead are not tagged as a pun-rep culture based on our selected keywords. The description of their society fits the coop-rep equilibrium in our model:

"At the community level, social control is maintained through socialization from childhood into norms of proper conduct and through use of informal sanctions such as gossip or ostracism. Individuals seek to avoid the embarrassment or shame of improper behavior, as well as to earn religious merit by following the major Buddhist rules of conduct (do not lie, steal, drink alcoholic beverages, fornicate, or kill living creatures). Certain kinds of misbehavior are thought to bring punishment from supernatural beings, usually in the form of illness. Although police and law courts exist, many people avoid using them except when absolutely necessary."

For the analysis of conflict, we match cultures in eHRAF to groups in the Ethnographic Atlas using the correspondence table provided by eHRAF.[13] For the analysis of ESS data, we aggregate the information on ethnic groups to the country level, following a procedure similar to Alesina et al. (2013) and Giuliano and Nunn (2018). First, we use the match created by Giuliano and Nunn (2018) between the Ethnographic Atlas and the *Ethnologue: Languages of the World* (Gordon and Grimes, 2009) dataset, to assign the revenge indicator to modern-day language groups. Giuliano and Nunn (2018) provide a shapefile that associates each of the approximately 7,500 language groups in the Ethnologue to a polygon; polygons are non-overlapping and mutually exclusive. Next, we intersect this shapefile with a shapefile of country borders and overlay the resulting polygons of ethnic groups in a specific country with a raster file of the world's population taken from *LandScan Global Population Database* (LandScan 2016). The LandScan raster reports population estimates for each 30 arc-second by 30 arc-second cell. By summing up all cells within each group-country-specific polygon,

---

[13]The table can be found at https://hraf.yale.edu/resources/reference/ea-cases-in-ehraf/.

we compute the total population of an ethnic group in a country. Finally, we aggregate the eHRAF revenge indicator at the country level, by taking the average of revenge indicators across all ethnic-groups with non-missing values, using the groups' shares of the country's total population as weights.

B.2   Construction of conflict dataset

Data on conflict between the government and groups not in power comes from the *Ethnic Power Relations Dataset Family* (EPR) (Vogt et al., 2015). The EPR tracks politically relevant groups for all countries in the world and for every year between 1946 and 2021. Groups are defined as politically relevant if they hold government power, or are officially excluded from power or discriminated. The status of groups in government is listed as "dominant", "monopoly," "senior partner" or "junior partner." Groups outside the government are classified as "discriminated", "powerless' or "self-excluded." In cases where a senior partner and one or several junior partners jointly govern, we use the senior partner to classify the government.[14] For each country and year, we then have a set of pairs between the government and each other politically relevant group.

Data on conflict comes from the *Uppsala Conflict Data Program* (UCDP). The UCDP is a dataset of geo-referenced conflict incidents with at least 25 fatalities between 1989 and 2016, with information on involved actors. Actors in the UCDP are matched to groups in the EPR through the ACD2EPR dataset. Our main dependent variable is a conflict indicator that takes on the value one if the government is in conflict with another group in a given year, and zero otherwise.

Ethnic groups in the EPR are linked to variables from the folklore dataset and other ethnographic controls through the Ethnographic Atlas. Michalopoulos and Papaioannou (2016) provide a concordance table between EPR and the Ethnographic Atlas for Africa. We extend this match to Asia by combining spatial and language information on groups from various sources. Information on groups' location comes from the GeoEPR dataset (Wucherpfennig et al., 2011), which assigns groups to spa-

---

[14]A procedure like the one followed in Guarnieri (2022), that aggregates characteristics of groups in multi-group governments and weights senior partners more than junior partners yields very similar results. Available upon request.

tial polygons; information on language comes from from the EPR Ethnic Dimensions (EPR-ED) Dataset (Bormann et al., 2017). Using the coordinates on groups' ethnic homelands and information on language provided by the Ethnographic Atlas, we identify groups in the Ethnographic Atlas that inhabit the polygons of groups in the EPR and speak the same language. We consider those groups as matched.[15] For remaining groups, we perform a manual match based on online searches.

B.3   Variable definitions and sources

Folklore motifs

Distance revenge motifs. Absolute difference between ethnic groups or countries in the number of revenge-related motifs as a proportion of total motifs, from Michalopoulos and Xue (2021). Revenge-related motifs are those tagged by the following keywords: *revenge, punish, punishment, penalty, retaliate, retaliation.*

Distance violence motifs. Absolute difference between ethnic groups or countries in the number of violence-related motifs as a proportion of total motifs, from Michalopoulos and Xue (2021). Violence-related motifs are those tagged by the following keywords: *violence, violent, battle, fighting, attach, soldier, guard, troop, army, enemy, fighter, invasion, invade, defender.*

Distance cheating motifs. Absolute difference between ethnic groups or countries in the number of deception-related motifs as a proportion of total motifs, from Michalopoulos and Xue (2021). Deception-related motifs are those tagged by the following keywords: *cheat, deceive, trick.*

---

[15]In this step, we are also aided in finding a correspondence between the Ethnographic Atlas and the EPR dataset by the A-MAR project (Birnir et al., 2018). The Ethnographic Atlas contains groups that are often aggregated into "upper level" groups in the EPR. The A-MAR project makes distinctions between "upper" and "lower" level groups, allowing us to assign Ethnographic Atlas groups to more aggregated groups that can then be matched to the EPR. For example, according to A-MAR the Semang are included among the Orang Asli, which appear in the EPR.

Country-level variables

Contiguous. Indicator for contiguous host-origin country pairs, from Guiso et al. (2009).

Common official language. Indicator for common official or primary language between host-origin country pairs, from Guiso et al. (2009).

Former colonial relation. Indicator for host-origin country pairs that have ever been in a colonial relationship, from Guiso et al. (2009).

Distance between capitals. Distance (as the crow flies) between host-origin country capitals in km, from Guiso et al. (2009).

Difference log GDP per capita. Absolute difference in log GDP per capita in 2000. GDP data is from the World Bank.

Religious similarity. Measure from Guiso et al. (2009). Product of the fraction of individuals in country j and in country i who have religion k, summed across all religions. Data on religious adherence is from the World Values Survey.

Distance patience. Absolute difference between host and origin country in patience from the GPS (Falk et al., 2018). The country-level measure of patience is a weighted average of a self-assessment of the willingness to wait and an intertemporal choice sequence using staircase method.

Distance risk-taking. Absolute difference between host and origin country in risk-taking preference from the GPS (Falk et al., 2018). The country-level measure of risk-taking is a weighted average of a self-assessment of the willingness to take risks in general and a lottery choice sequence using staircase method.

Distance positive reciprocity. Absolute difference between host and origin country in positive reciprocity from the GPS (Falk et al., 2018). The country-level measure of positive reciprocity is a weighted average of a self-assessment of the willigness to return a favor and agreement with a statement on gift-giving in exchange for help.

Distance negative reciprocity. Absolute difference between host and origin country in negative reciprocity from the GPS (Falk et al., 2018). The country-level measure of negative reciprocity is a weighted average of three self-assessment questions on the willingness to take revenge, the willingness to punish unfair behavior toward one's self and the willingness to punish unfair behavior toward others.

Distance altruism. Absolute difference between host and origin country in altruism from the GPS (Falk et al., 2018). The country-level measure of altruism is a weighted average between a donation decision and a self-assessment of the willingness to give to good causes.

Distance trust. Absolute difference between host and origin country in trust from the GPS (Falk et al., 2018). The country-level measure of trust is based on agreement with the statement that people have only the best intentions.

Conflict

Conflict incidence. Pair-level indicator for conflict in a given year between the government and another politically relevant group. Based on data from the the ACD2EPR dataset (Vogt et al., 2015).

Ethnic group-level variables

Linguistic distance. We compute cladistic linguistic distance following Fearon (2003). First, we merge the language of each ethnic group to the Ethnologue. The Ethnologue includes information on each language's position in a linguistic tree that starts with a broad language family and branches out to intermediate nodes before ending with the language itself. Distance between a pair of languages i and j is computed as

$$d_{i,j} = 1 - \left(\frac{\#\text{of common nodes between } i \text{ and } j}{\frac{1}{2}(\text{of nodes of language } i + \text{of nodes of language } j)} \right)^{\lambda}$$

where λ is a concavity factor used to attribute higher weight to earlier common nodes, as separation in earlier nodes implies higher linguistic divergence. We set $\lambda = 0.5$

following existing literature (Fearon and Laitin, 2003; Guarnieri and Tur-Prats, 2020; Guarnieri, 2022).

Religious distance. We implement the same formula as above to compute religious distance following Guarnieri and Tur-Prats (2020). We use religions reported for each ethnic group in the EPR-ED dataset (Bormann et al., 2017).

Geographic distance. Simple geodesic distance between centroids of group polygons, from GeoEPR (Wucherpfennig et al., 2011).

Distance jurisdictional hierarchy. Absolute difference between ethnic groups in variable *v33* of the Ethnographic Atlas, capturing jurisdictional hierarchy beyond the local community. The variable takes on values between 1 (no hierarchy beyond local community) and 5 (four levels of hierarchy).

Distance settlement complexity. Absolute difference between ethnic groups in variable *v30* of the Ethnographic Atlas. The variable takes on values from 1 (nomadic or fully migratory) to 8 (complex settlements), with higher values indicating more permanent and complex settlements.

Distance High Gods. Absolute difference between ethnic groups in the presence of High Gods, based on *v34* of the Ethnographic Atlas. We code High Gods as present if the variable takes on values between 3 (Not active in human affairs) and 5 (Supportive of human morality). Results are similar when we code the presence of High Gods only when High Gods are moralizing ($v34 = 5$).

Distance herding dependence. Absolute distance between ethnic groups in dependence on herding. We code this variable for each ethnic group following Becker (2021) and Cao et al. (2021). We use the variable *v4* of the Ethnographic Atlas, which codes societies by levels of dependence on animal husbandry, ranging from 0-5% to 86-100%. We recode this to the median of each category and multiply with a dummy that equals one if the predominant animal is a herding animal (sheep, cattle, horses, reindeer, alpacas, or camels, based on *v40* of the Ethnographic Atlas).

Distance segmentary lineage. Absolute difference between ethnic groups in an indicator for segmentary lineages. This variable comes from Moscona et al. (2019) and is only available for 145 ethnic groups in Africa.

Distance agricultural suitability. Absolute distance between ethnic groups in average suitability for low-input rainfed agriculture in the ethnic groups' territories. Suitability data is from FAO (Fischer et al., 2002), averaged over polygons in the GeoEPR dataset (Wucherpfennig et al., 2011).

Distance ruggedness. Absolute distance between ethnic groups in average ruggedness in the ethnic groups' territories. We compute terrain ruggedness following Nunn and Puga (2012).Underlying elevation data are from *GTOPO30*, a global digital elevation model (DEM) of 30 arc-second by 30 arc-second cells. Ruggedness is averaged over polygons in the GeoEPR dataset (Wucherpfennig et al., 2011).

Distance elevation. Absolute distance between ethnic groups in average elevation in the ethnic groups' territories. Data on elevation is provided by EPR (Vogt et al., 2015).

Distance group size. Absolute distance in group size between ethnic groups. Information on group size is provided by EPR (Vogt et al., 2015).