

War and Relatedness*

Enrico Spolaore
Tufts University and NBER

Romain Wacziarg
UCLA and NBER

August 2015

Abstract

We examine the empirical relationship between the occurrence of interstate conflicts and the degree of relatedness between countries, measured by genetic distance. We find that more closely related populations are more prone to go to war with each other, even after controlling for numerous measures of geographic distance and other factors that affect conflict. We provide an economic interpretation of these findings, based on two connected mechanisms. More closely related groups tend to have more similar preferences over rival goods, such as natural endowments, and are thus more likely to fight over them. In addition, rulers have stronger incentives to engage in conflict over territories inhabited by populations that are more similar to their own homeland populations, to minimize the political costs of ruling over people with different views over government types and policies. These mechanisms are supported by empirical evidence on international conflicts over natural endowments and on territorial changes, including those associated with the formation of new states and decolonization.

*Spolaore: Department of Economics, Tufts University, Medford, MA 02155-6722, enrico.spolaore@tufts.edu.
Wacziarg: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles CA 90095, wacziarg@ucla.edu.
We are grateful to Quamrul Ashraf, Don Cox, Klaus Desmet, Rajeev Dehejia, Quy-Toan Do, Steven Durlauf, Jim Fearon, Jeff Frieden, Oded Galor, Michelle Garfinkel, Tim Groseclose, Gordon Hanson, Yannis Ioannides, Michael Klein, Glenn Loury, Philippe Martin, Deborah Menegotto, Massimo Morelli, Ömer Özak, Daniele Paserman, Ken Shepsle, Vickie Sullivan, three anonymous referees, and participants at numerous seminars and conferences for helpful comments.

1 Introduction

Is war more likely between states that differ in ancestry, ethnicity, and historical legacies? Many argue that there is a general tendency towards violent confrontation between ethnically distant groups. For example, Bremer (2000, p. 27) wondered whether "cultural differences [...] should lead to misunderstandings, stereotyping, clashes of values, and so forth, which in turn promote intercultural fights." This debate can partly be traced back to the sociologist William G. Sumner (1906), who formulated the primordialist view that ethnic dissimilarity is associated with war and plunder, while societies that are ethnically and culturally related tend to fight less with each other.

This paper presents new empirical findings on the determinants of conflict between sovereign states, strongly supporting the opposite view. States whose populations are *more* closely related are significantly more likely to engage in international conflict with each other, even after controlling for a variety of measures of geographic proximity and other factors affecting conflict.

Why are more closely related states more likely to engage in international wars? Our explanation is based on two connected mechanisms, stemming from the fact that more closely related populations tend to be more similar along a range of historical and cultural traits.¹ First, more closely related populations tend to have more similar preferences over rival goods, such as natural endowments or historic sites, and are thus more likely to fight over them. This mechanism directly translates into a higher likelihood of international conflict between related populations over the control of territories and natural resources with given characteristics - for example, fertile soil in an agricultural society or oil in more recent times. Second, more closely related people on average also share more similar preferences over policies and types of government. Therefore, governments in control of more similar populations bear lower political costs, and face a lower probability of internal conflict and rebellion. Rulers then have stronger incentives to engage in conflict over territories inhabited by people that are historically and culturally more similar to their own homeland populations. In sum, the effects of relatedness on preferences over both rival goods and government policies imply that states inhabited by more closely related citizens face stronger incentives to fight with each other over the control of the same resources, territories, and populations.

To measure historical relatedness between populations, we use genetic distance. Genetic distance measures the difference in gene distributions between two populations, where the genes under consid-

¹In the working paper version of this study, we presented a model of relatedness and conflict formalizing these ideas (Spolaore and Wacziarg, 2013a).

eration change randomly and independently of selection pressure. Most random genetic change takes place regularly over time, acting as a molecular clock. Consequently, genetic distance captures the time since two populations have shared common ancestors, i.e. since they were the same population. Divergence in neutral genes provides information about lines of descent, so that genetic distance is a summary measure of general relatedness between populations.² Heuristically, the concept is analogous to relatedness between individuals: two siblings are more closely related than two cousins because they share more recent common ancestors: their parents rather than their grandparents. Since many characteristics, including cultural traits and preferences, are transmitted across generations over the long run, genetic distance provides a comprehensive measure of distance in such traits across populations. Indeed we document empirically that genetic distance is correlated with a wide set of cultural differences.

This paper's main empirical result is that populations that are more closely related are more likely to engage in interstate conflict and wars, and this effect is substantial in magnitude. A one-standard-deviation increase in genetic distance between two countries reduces that pair's probability of ever being in conflict between 1816 and 2001 by 23.84%. The effect of genetic distance is even higher (36.79%) when we instrument for modern genetic distance using genetic distance between populations as of the year 1500, to account for measurement error and potential endogeneity issues.

The negative effect of genetic distance holds when controlling for measures of geography, such as contiguity, geodesic distance, latitudinal and longitudinal differences, and other measures of geographic barriers and isolation, including differences in terrain ruggedness and a measure of travel time in the pre-industrial era. Once genetic distance is taken into account, geographic variables have smaller effects, although they remain significant. This suggests that the paramount effect on conflict attributed by some scholars to geographic proximity (e.g. Gleditsch and Singer, 1975, Vasquez, 1995) is not entirely causal, but partly due to the fact that neighboring populations are culturally and historically related. In addition, the effect of genetic distance is robust when accounting for differences in military capabilities and income per capita across countries.

We also consider the effects of direct measures of cultural distance, such as religious and linguistic distance, finding that these measures decrease the probability of conflict: countries that are linguistically and religiously closer on average fight more with each other. The effects of linguistic distance

²Our measures of genetic distance are fixation indices (F_{ST}) between human populations from Cavalli-Sforza et al. (1994). The F_{ST} index was first suggested by the geneticist Sewall Wright, who, coincidentally, was the brother of Quincy Wright, who pioneered empirical research on conflict (Wright, 1942).

and religious distance on international conflict are in line with our interpretation of the mechanisms at work. For instance, we expect that states sharing similar religions would engage in more conflict over control of coreligionist populations and common holy sites, such as Jerusalem for Abrahamic religions.³ The fact that religious distance reduces the likelihood of war would be hard to rationalize within a clash-of-civilizations view (Huntington, 1993).

At the center of our interpretation of the empirical evidence is the hypothesis that more closely related populations share more similar preferences over rival and excludable goods. Direct tests of this mechanism are provided by evidence on conflict about specific rival goods. A prominent example is oil, which has played a key role in international relations and confrontations in recent decades. The likelihood that two countries would go to war with each other after 1945 is higher when there is oil in their territories. Even more importantly, the effect of oil on conflict is much higher for countries that are more closely related (that is, which are at a smaller genetic distance from each other), again controlling for a vast range of geographical variables. This is an instance of the general mechanism we have emphasized: more closely related populations go to war with each other because of closer preferences over rival goods, including specific natural resources. In addition, we consider other measures of endowments, soil quality and the extent of territory with a temperate climate, which would be more valuable for agricultural economies. The presence of these endowments is associated with a higher likelihood of conflict in our sample for the earlier period (i.e., before 1900), and the effect is larger for countries that are genetically closer.

Further evidence in support of our interpretation of the relation between relatedness and conflict is provided by historical data on territorial change between 1816 and 2008. As expected, we find that violent territorial change on average occurred between more closely related states. Consistent with our interpretation, we find a strong effect of relatedness when the territorial change is associated with the formation of a new state. In those cases, violent conflict occurred when the secessionists and the central government were more closely related, and therefore the central government had more to lose from letting the breakaway region go. In contrast, peaceful transitions to independence took place for populations that were breaking away from more distantly related central governments. These effects played a key role in the process of decolonization: overall, this process has been mostly peaceful, consistent with the pacifying effects of relatively low relatedness between colonizers and former colonies. The few instances of violent decolonization occurred when populations in former colonies

³Another example is the site of Prince Vladimir's baptism in Crimea for Orthodox Christians. For a recent discussion of the role of "sacred land" in the conflict over Crimea, see Turchin (2014).

were culturally closer to those in the colonizer country. For instance the American Revolution, started by British subjects against their central government, was bloody, but most British colonies in Africa gained their independence without wars during the 20th century.

This paper is part of a small but growing empirical literature on the connections between long-term relatedness and societal outcomes (for instance Spolaore and Wacziarg, 2009).⁴ More broadly, this study is related to the evolutionary literature on cultural transmission of traits and preferences (e.g., Bisin and Verdier, 2001, Richerson and Boyd, 2004) and to the growing empirical literature on the deep historical roots of economic and political outcomes. Recent contributions include Ashraf and Galor (2013) and Putterman and Weil (2010). For an overview, see Spolaore and Wacziarg (2013b). Our paper is also part of a growing economics literature on conflict and international wars, pioneered by Schelling (1960) and Boulding (1962), and surveyed by Garfinkel and Skaperdas (2007). Recent empirical contributions include Martin, Mayer and Thoenig (2008) and Caselli, Morelli and Rohner (2015).

2 Data and Methodology

2.1 Measuring Conflict

We use panel data on interstate conflict between 1816 and 2001 from the Correlates of War Project (www.correlatesofwar.org).⁵ We start from a discrete indicator of the intensity of a bilateral conflict between countries i and j in year t . In any given year, the indicator takes on a value from 0 for no militarized conflict to 5 for an interstate war involving more than 1,000 total battle deaths. Following the convention in the literature, we define a dummy variable C_{ijt} equal to 1 if the intensity

⁴There exists a small theoretical and empirical literature on war and evolution. Contributions by economists are Hirshleifer (1998), who discussed the evolutionary motives for warfare, and Bowles (2009), who studied whether warfare between ancestral hunters-gathers affected the evolution of group-beneficial behavior. More recently, Arbatli, Ashraf and Galor (2015) use genetic diversity within each country and study civil conflict. They find that genetically more diverse populations are more likely to engage in civil conflict. Their finding is not inconsistent with our results, as it is about conflict within countries, not across countries. On the contrary, their result is consistent with the hypothesis that more similar populations may share more similar preferences over common policies and government types, which also implies that rulers of different countries may face stronger incentives to conquer territories inhabited by peoples more similar to those they already rule.

⁵See Jones et. al. (1996) and Faten et al. (2004) for details concerning the coding of bilateral militarized disputes in the Correlates of War database.

of militarized conflict is equal to or greater than 3, zero otherwise. Since our main independent variable is time-invariant, our focus is mainly cross-sectional. Thus, we look for pairs that were ever involved in a conflict ($C_{ijt} = 1$ for some t) over the time period 1816-2001: the pair is coded as having experienced a conflict during this period if there was a conflict in at least one year. Our main dependent variable is this binary indicator of conflict, denoted C_{ij} . We separately examine the determinants of full-blown war (corresponding to a pair having ever experienced a conflict intensity equal to 5), as well as those of the maximal intensity of conflict.⁶

2.2 Measuring Relatedness

To capture relatedness between populations, we use genetic distance. Since the interpretation and construction of this measure was discussed in detail in Spolaore and Wacziarg (2009), we provide a shorter description here. The measure of genetic distance that we use, F_{ST} , is a summary measure of differences in allele frequencies across a range of neutral genes, and captures the length of time since two populations became separated from each other. When two populations split apart, random genetic mutations result in genetic differentiation over time. The longer the separation time, the greater the genetic distance computed from a set of neutral genes. In other words, F_{ST} genetic distance is a direct measure of relatedness between populations. The specific source for our data is Cavalli-Sforza et al. (1994), pp. 75-76.⁷

Our focus is on a set of 42 world populations for which there is data on bilateral genetic distance, computed from 120 neutral alleles (examples include genes that affect blood types). Among the set of 42 world populations, the maximum genetic distance is between Mbuti Pygmies and Papua New-Guineans ($F_{ST} = 0.457$), and the minimum is between the Danish and the English ($F_{ST} = 0.002$). The mean genetic distance among the 861 available pairs is 0.134.

While the data on genetic distance is available at the level of populations, the rest of our data is at the country-pair level. It was therefore necessary to match genetic groups to countries. The

⁶In order to appear in the Correlates of War dataset, a country needs to be a sovereign state. To appear in our sample, we need, in addition, to have available data on control variables. Most countries in our sample were continuously in existence since 1816. To the extent that some countries, like Germany and Italy, were formed as sovereign states during the 1816-2001 period, C_{ijt} is coded only for the years when they are in existence. Thus, C_{ij} for these countries is defined only over the relevant years. Due to lack of data on control variables, countries that ceased to be sovereign states (like the Kingdom of Bavaria or the Grand Duchy of Tuscany) do not appear in our sample at any time.

⁷Our results are robust to using alternative measures of genetic distance, such as Nei's distance. The corresponding estimates are available in the online appendix.

procedure to match populations to countries is described in detail in Spolaore and Wacziarg (2009). To summarize, each of the 42 groups was matched to almost all of the 1,120 ethnic groups in Alesina et al. (2003). The same source provided the distribution of these ethnic groups across virtually all the countries in the world. Thus, we could construct measures of genetic distance between countries, rather than groups. We constructed two such measures. The first was the distance between the plurality ethnic groups of each country in a pair, i.e. the groups with the largest shares of each country’s population. The second was a measure of weighted genetic distance, constructed as follows: assume that country i is composed of populations $m = 1 \dots M$ and country j is composed of populations $n = 1 \dots N$.⁸ Denote by s_{im} the share of population m in country i (similarly for country j) and d_{mn} the genetic distance between populations m and n . The weighted F_{ST} genetic distance between countries i and j is then:

$$FST_{ij}^W = \sum_{m=1}^M \sum_{n=1}^N (s_{im} \times s_{jn} \times d_{mn}) \quad (1)$$

where d_{mn} is the F_{ST} genetic distance between groups m and n , and s_{km} is the share of group m in country k (group shares from Alesina et al., 2003, usually pertain to dates in the 1990s, with a few as early as 1983 and some as late as 2001). This measure represents the expected genetic distance between two randomly selected individuals, one from each country. Weighted genetic distance is very highly correlated with genetic distance based on plurality groups (the correlation is 93.2%). We use the weighted F_{ST} distance as the baseline measure throughout this study, as it is a more precise measure of average genetic distance between countries, but the results are robust to using genetic distance between plurality groups instead. The corresponding estimates are available in the online appendix.

The match of populations to countries pertains to the contemporary period, after the great migrations that followed the European conquest of the New World. Hence, for instance, the plurality population in Australia for the current period is the English population. We also matched countries to their populations in the year 1500, before modern explorations and migrations. For example, for 1500 Australia is matched to the Australian Aborigines rather than the English. Genetic distance between countries using the 1500 match can be used as an instrument for current genetic distance.

⁸That is, we do not treat countries formed by different ethnic groups as a new population, in a genetic sense, but as a set of separate populations. This is consistent with the idea that different groups have inherited different traits and preferences from their ancestors, and that the country’s traits and preferences are a weighted average of the traits and preferences inherited by the different groups.

Since we do not have detailed data on ethnic composition in 1500, the corresponding match only refers to plurality groups. Matching countries to populations for 1500 is more straightforward than for the current period, because Cavalli-Sforza et al. (1994) attempted to sample populations as they were in 1500. The correlation between the two measures (weighted genetic distance based on current populations and genetic distance based on plurality groups as of 1500) is 0.723 in our baseline sample.

We hypothesize that genetic distance is associated with differences in a broad vector of traits transmitted intergenerationally with variations. The longer the separation times between countries or populations, the greater the average distance in these traits. In particular, these include preferences over rival goods, as well as, more broadly cultural values and other intergenerationally transmitted characteristics. The relationship between genetic distance and cultural distance has been studied by Desmet et al. (2011), who found a strong association between genetic distance and a metric of distance in answers to questions from the World Values Survey (WVS) for a sample of European populations. We extended this investigation to a worldwide sample by computing indices of the average distance in answers to 98 questions from the WVS integrated questionnaire (covering all waves of the survey), available for 73 countries (2,628 country pairs). The choice of questions was dictated by the availability of answers for a sufficient number of countries. For each question, we computed the Euclidian distance in the average answer shares of each option.⁹ These distances were normalized to have mean 0 and standard deviation 1, so as to bear equal weights in indices that aggregate them. We computed several such indices. First, we calculated the average distance across all 98 questions. Then, we broke down the questions by each of the 7 question categories provided by the WVS and computed average distance for these subsets. Finally, we computed separate distance metrics for binary and non-binary questions. We then investigated the relationship between these indices of cultural distance and our preferred weighted measure of F_{ST} genetic distance.

We found positive and usually sizeable correlations between cultural distance and genetic distance. The most aggregated measure, across all 98 questions, bears a correlation of 0.27 with genetic distance, based on the 2,628 observations. By question category, the strongest correlations are with categories A (Perceptions of Life, correlation: 0.25), E (Politics and Society, correlation: 0.30) and F (Religion and Morale, correlation: 0.22). The correlation with genetic distance is also stronger for questions with more than two possible answers (correlation: 0.32) than those with binary answers

⁹There were 35 binary questions where the Euclidian distance is just the absolute value of the difference in the answer shares to one of the two possible responses. For the remaining 63 questions with more than two possible answers, the Euclidian distance is the square root of the sum of squared differences in response shares.

(correlation: 0.15). In regression analysis controlling for a wide array of geographic distance measures, we confirmed a highly significant positive relationship between genetic distance and our WVS-based cultural distance measures. Overall, these findings lend support to the use of genetic distance as a summary measure of a broad array of intergenerationally transmitted traits reflecting preferences, values, norms and culture (for further evidence along these lines, see Spolaore and Wacziarg, 2015).

2.3 Summary Statistics

Table 1 and 2 provide basic statistics that give a general sense of patterns in the data. The statistics pertain to a baseline sample of 13,575 country pairs, based on 176 underlying countries. Table 1, Panel A provides means and standard deviations. Conflict is a relatively rare phenomenon, as only 5.6% of country pairs ever experienced a conflict between 1816 and 2001. War is even more rare, with an incidence of 2.1%. Panel B provides pairwise correlations between the main variables in the analysis. We observe a negative correlation between genetic distance and both conflicts and wars, and the other correlations are of the expected size and magnitude.

Table 2 shows the conditional frequency of both wars and conflicts, confirming that wars are rare occurrences, as only 275 country pairs out of 13,175 pairs have ever experienced full-blown wars between 1816 and 2001. Almost 28% of these wars occurred between countries in the bottom decile of genetic distance, and almost 54% of all wars occurred in pairs in the bottom quartile. Only 10 wars were observed in pairs in the top quartile, of which 7 involved South Africa as one of the combatants.¹⁰ In sum, countries that are very genetically distant almost never went to war with each other in our sample. The same statements hold when conditioning on measures of geographic distance, as is also done in Table 2: even wars occurring across large geographic distances typically involve mostly genetically similar participants. For instance over half of the wars occurring between non-contiguous countries involved country pairs in the bottom quartile of genetic distance.

Similar observations hold when we consider more broadly militarized conflicts rather than only wars. While there are vastly more pairs that were ever involved in such conflicts (744 versus 275), the relative frequency by quartile of genetic distance is roughly preserved. Similarly, the proportions do not change much when conditioning on geographic distance being large between the countries in

¹⁰While South Africa is characterized in our data as genetically distant from European populations due to its large African majority, a historical examination of wars involving South Africa reveals that the wars were spurred mainly by conflicts over issues separating European powers and South Africa's European power elite. Thus, in this instance genetic distance is coded in a way that works *against* finding a positive link between relatedness and conflict.

a pair - countries not sharing a common sea or ocean, non-contiguous countries, or countries that are more than 1,000 kilometers apart. Thus, Table 2 provides suggestive evidence that relatedness and conflict are positively related. To examine this hypothesis more formally we turn to regression analysis.

2.4 Empirical Specification

A regression setup allows us to control for various determinants of interstate militarized conflicts, in particular a range of geographic distance metrics. As a starting point for our empirical specification, we follow the practice in the existing literature (for instance Bremer, 1992, Martin, Mayer and Thoenig, 2008), regressing a binary indicator of interstate conflict on a set of bilateral determinants. Our approach is cross-sectional. Since our main independent variable of interest, F_{ST} genetic distance, is time invariant at the horizon of this study, a natural starting point is to consider the determinants of whether a country ever had a conflict or a war between 1816 and 2001. The baseline specification is:

$$C_{ij} = \beta_1 X_{ij} + \beta_2 F_{ST_{ij}}^W + \eta_{ij} \quad (2)$$

where the vector X_{ij} contains a series of time-invariant controls such as a contiguity dummy, log geodesic distance, log longitudinal and latitudinal distance, several other indicators of geographic isolation, and dummy variables indicating whether the countries in a pair were ever part of the same polity and were ever in a colonial relationship.

Equation (2) is estimated using probit. Throughout, we report marginal effects evaluated at the mean of the independent variables, and report the standardized magnitude of the effect of genetic distance (the effect of a one standard deviation change in genetic distance as a percentage of the mean probability of conflict). Because the proportion of pair-year observations with conflicts is small, in order to improve the readability of the marginal effects we multiplied all of them by 100 in all tables. The proper interpretation of the estimates displayed in the tables, then, is as the marginal effect of each variable on the probability of conflict in percentage points.

2.5 Discussion of the Identifying Assumptions

Our empirical goal is to estimate the determinants of interstate wars. Therefore, by definition, our units of observation are not groups or populations, but sovereign states. Our dependent variable is not violent conflict between any pairs of groups or populations in general, but only international

conflict - i.e., conflict between groups that have formed separate states. These facts have important implications for the interpretation of our results and for our identification strategy.

A negative effect of genetic distance on international conflict should not be interpreted as a general claim that any randomly selected pair of groups that are more closely related should necessarily face a higher probability of engaging in violent conflict with each other, irrespective of their political organization. In fact, the logic of our own analysis implies that genetically closer populations have at least two good reasons to want to form a unified state. One reason is that closely related populations tend to have more similar preferences over types of government and common policies. Hence, they face lower political costs when forming a unified state, as emphasized by the political-economy literature on state formation (e.g., Alesina and Spolaore, 1997, 2003, 2006; Alesina, Spolaore and Wacziarg, 2000, Desmet, LeBreton, Ortuño-Ortín and Weber, 2010). Another reason stems from similarity of preferences over rival goods. More similar populations are more likely to want the same rival goods, and hence more likely to fight over them in an anarchic world. Thus, they have stronger incentives to set up institutions - such as sovereign states - that establish and enforce property rights over private goods, thus reducing the opportunities for violent conflict among their citizens. In sum, populations that are genetically close will often be part of the same country, and therefore, by definition, will not engage in international conflict. That still leaves open the question of whether and why, among groups that are not part of the same country, conflict is more prevalent among genetically closer populations. This is the question addressed in this paper.

The goal of our empirical analysis is to estimate the effect of relatedness on international conflict. In order to address potential endogeneity concerns, it is useful to consider the sources of geographical variation in genetic distance, and the extent to which such variation is the result of forces that have operated independently of the process of modern nation state formation and modern conflict. Here, we must draw a distinction between the Old World (Eurasia and Africa) and the New World (the Americas and Oceania). In the Old World, the current geographic distribution of genetic distance is predominantly the outcome of forces that operated in pre-modern times (i.e. before 1500 CE), before the process of modern nation-state formation. In fact, according to the scientific literature (e.g., Cavalli-Sforza, Menozzi and Piazza, 1994, Bellwood, 2013), the geographic variation in genetic distance before 1500 was determined to a large extent by prehistorical patterns of human migration. Our *Homo sapiens* ancestors migrated out of East Africa and progressively populated Eurasia and Oceania starting approximately 70,000 years ago, while later making their way across the Bering straits and into the Americas at the end of the last glacial period, about 15,000 years ago. Along these

migratory routes, groups of humans separated from each other. Genetic distance in the Old World is very closely related to these separation times, which are precisely what genetic distance is designed to capture. It is important to notice that the determinants of such separations and locations in prehistoric and pre-modern times were mostly geographical and biogeographical forces (demographic pressure, extinction of prey, ice-age shocks, and so on) affecting hunters-gathers or early agriculturists or pastoralists who had not yet formed modern nation states. In contrast, the current distribution of genetic distance in the New World is the outcome of large-scale migrations from Europe and Africa, and intermixing of populations, which are closely associated to the formation of modern states and colonial empires. In our empirical analysis, we exploit the pre-modern geographic distribution of genetic distance - i.e., genetic distance around 1500, before the formation of modern nation states and colonial empires - to identify the effects of modern genetic distance on modern wars.

We use genetic distance in 1500 as an instrument for contemporary genetic distance. The main goal is to address the specific concern regarding endogeneity that stems from population movements since the discovery and colonization of the New World. To the extent that past conquests triggered movements of populations between countries, and to the extent that past conflicts are conducive to a higher propensity for current conflict, country pairs could have a lower genetic distance because of their high (past and present) propensity to enter into militarized conflicts. This would lead to an upward bias (in absolute value) in estimates of the effect of genetic distance. Genetic distance in 1500 is unlikely to be causally affected by conflicts between 1816 and 2001, and it is unlikely to affect contemporary conflict through channels other than its effect on modern genetic distance, so that it satisfies the conditions for a valid instrument.¹¹ Another concern addressed by the use of genetic distance in 1500 as an instrument is measurement error, stemming from imperfect matches of genetic groups to current populations and countries, leading to probable attenuation bias. Matching countries to genetic groups is much more straightforward for 1500 for two reasons. First, Cavalli-Sforza et al. (1994) explicitly collected data for populations as they were around 1500 - that is, they took care to sample only direct descendants of aboriginal populations that had lived continuously at that location since 1500, not people whose ancestors had moved to the current location after the great migrations post-1500. Second, matching genetic groups to countries is easier for the period

¹¹Population geneticists have noted that a very high degree of admixture from migration or conquest would be required in order to significantly affect a country's genetic distance to others, so this endogeneity concern applies mostly to the colonization of the New World and movements of European settlers and African slaves that characterized it (Cavalli-Sforza, Menozzi, and Piazza, 1994).

predating the great migrations that followed the discovery of the New World, because there is no need to track the Old World origin of current New World populations. Finally, using 1500 genetic distance as an instrument is a way to address any remaining omitted variables concerns, beyond the inclusion of our wide set of control variables directly in the conflict specification, to the extent that 1500 genetic distance is uncorrelated with these omitted factors conditional on contemporary genetic distance.

3 Empirical Results

3.1 Baseline Estimates

Table 3 presents baseline estimates of the coefficients in equation (2). We start with a univariate regression (column 1), showing a negative relationship between genetic distance and the incidence of militarized conflict. The magnitude of this effect is large: one standard deviation change in genetic distance (0.068) is associated with a 68.81% decline in the percentage probability of two countries ever having experienced a conflict (in the cross-section, that baseline probability is 5.65% for the entire period between 1816 and 2001). Obviously, this estimate is tainted by omitted variables bias, stemming mainly from the omission of geographic factors correlated with both conflict and genetic distance.

Column (2), our baseline specification, introduces eight measures of geographic distance, plus two measures of colonial past. The choice of the geographic controls was motivated by the goal of controlling for dimensions of geographic distance that constitute barriers to militarized conflict. Contiguity is an obvious example, since two contiguous countries do not have to project force very far in order to fight each other, and might have adjacent territories under dispute. Access to a common sea or ocean facilitates conflict through the projection of a naval force. Geodesic distance, on the other hand, limits the ability to project force. A landlocked country may be harder to attack by a non-contiguous neighbor, since its armies would have to cross another country first. Finally, islands could be either more or less prone to conflict depending on whether surrounding seas afford protection from attack, whether this protection raises an island's propensity to attack others, or whether an island is easier to reach with ships (in fact, we find that pairs including islands are more prone to conflict). Empirically, these measures usually bear the expected signs (more distance, less conflict), and their inclusion reduces the effect of genetic distance.¹² However, this effect remains

¹²Proceeding sequentially, we found that adding these controls one by one progressively reduced the effect of genetic

negative and significant statistically. Its magnitude is substantial: a one standard deviation increase in genetic distance is associated with a reduction in the probability of conflict of 23.84% of that variable's mean.

In column 3, we address the possible endogeneity of genetic distance. The results using 1500 CE genetic distance as an IV reinforce those previously reported. Interestingly, the standardized effect of genetic distance rises by over 50% - to 36.79% - relative to the estimates of column (2), suggesting that the latter understated the effect. To adopt a conservative approach, in the rest of the analysis we provide estimates mostly without instrumenting, keeping in mind that non-instrumented probit estimates of the effect of genetic distance are possibly an understatement of its true magnitude.

The remaining columns of Table 3 consider the determinants of wars rather than conflicts more broadly (columns 4 and 5). Again, genetic distance reduces the propensity for war in a statistically significant way: a standard deviation increase in genetic distance reduces the probability of ever having experienced a war by 20.57% of this variable's mean, an effect comparable to that for conflict more broadly. As before, the standardized magnitude of the effect rises (here by about 40%) when instrumenting with genetic distance as of 1500.¹³

3.2 Further Geographic Controls

While our baseline estimates include a wide array of geographic controls, a concern remains that some omitted geographic factor may drive the result. In Table 4 we add an even longer list of geographic controls to the baseline specification, while retaining the baseline set of controls from Table 3.

In column 1 we add controls representing differences in the topographical features of the land areas of each country in a pair, obtained from Nunn and Puga (2012). The first of these variables is the distance, but that after adding four controls the estimated probit marginal effect of genetic distance stabilized around 20. The order did not matter much. The largest reductions in the coefficient on genetic distance were found for contiguity, log geodesic distance, the landlocked dummy and the log product of land area, after which additional controls did not meaningfully reduce the effect of genetic distance. This gives us some confidence that we are adequately controlling for geographic impediments to conflict (Oster, 2014).

¹³Additionally, we examined whether the effect of genetic distance differs by type of conflict, exploiting information available in the COW database on the type of dispute. Non-territorial issues include a desire to change the other country's regime or to change the other country's policies (Vasquez and Henahan, 2001). We defined a territorial conflict as one for which either country seeks a territorial revision either as the most or second most important rationale for the dispute. We found that the effect of genetic distance was negative and statistically significant for both territorial and non-territorial conflicts. These empirical results are available in the online appendix.

absolute difference in the terrain ruggedness index.¹⁴ The others are the absolute bilateral differences in the percentage of fertile soil, the percentage of tropical area (defined by the corresponding Köppen-Geiger climatic zone), the percentage of desert area, the percentage of a country's area within 100 kilometers of an ice free coast, and the average distance to the coast. The inclusion of these variables slightly reduces the number of observations, but does not greatly affect the standardized effect of genetic distance, now equal to 20%.

In column 2 we introduce a measure of climatic similarity of the land areas of the two countries that define a pair: this index ranges from 0 to 100 and represents the average absolute difference in the percentage of the countries' land areas that belong to each of the 12 Köppen-Geiger climatic zones. The effect of genetic distance barely changes compared to the baseline in Table 3, while climatic differences are unrelated to conflict.

In column 3 we control for the average difference in elevation between two countries in a pair. This is meant to capture barriers such as mountain ranges that might impede conflict between two countries. This time the effect of genetic distance rises a bit, but the difference in elevation does not matter. This result was unchanged when we added the interaction between elevation differences and contiguity, to capture the fact that elevation differences may only matter when countries are contiguous.

In column (4) we introduce an alternative measure of the bilateral distance between two countries from Özak (2013), the Human Mobility Index with Seafaring (HMIsea). This index is based on the travel time, by land or by sea, between any two locations on the globe along the optimal path in the pre-industrial era, so high values of the index represent higher costs of moving from one point to another. It takes into account various impediments to travel such as topographical barriers, climatic conditions between locations, and type of terrain. It is potentially more predictive of genetic distance than geodesic distance since it more accurately predicts how difficult it was to effectively reach another location prior to the advent of modern transportation technology, and thus may be a closer predictor of long-run separation times between populations. We find that a greater pre-industrial mobility costs are negatively associated with the probability of conflict, but that the inclusion of this variable does not affect the estimate on genetic distance.

¹⁴Nunn and Puga (2012) present four variants of this index. We used the population-weighted variant as it is the only one that came out as a significant determinant of conflict. Like the other 3, however, its inclusion in the regression does not affect the magnitude of the effect of genetic distance.

Finally, in the last column of Table 4, we introduce all the above variables jointly, finding only a modest reduction in the effect of genetic distance, its standardized effect on conflict going from -23.83% in the baseline specification to -21.19% in the specification with all the geographic controls.

3.3 Estimates Across Time and Space

Genetic distance is usually larger between continents than within continents. In estimates included in the online appendix, we assessed whether our results are robust to continent effects and hold within continents. We first added to our baseline specification a variable equal to 1 if the two countries in a pair are from the same continent, 0 otherwise. The effect of genetic distance barely changed. We next conducted a much more stringent test: for each of five continents we included two dummies, one equal to one if the two countries in a pair are from that continent, the other equal to one if one country in the pair is on that continent, but not the other. Jointly these 10 continent dummies accounted for 82% of the variation in genetic distance (this is the R^2 in a regression of weighted F_{ST} genetic distance on the 10 continent dummies, omitting the constant term). However, the inclusion of these dummies in the regression did not eliminate the effect of genetic distance on conflict.

To examine if the negative effect of relatedness on conflict differs across various geographic locations, we next broke down the sample by continent, defined as Europe, Asia, Africa and the Americas. This is a way to further address the possibly confounding effects of geographic factors, since geographic barriers to conflict tend to be larger across continents than within continents. We found a negative effect of genetic distance on conflict within every continent, with significant effects at the 1% level for Europe and at the 5% level for the Americas (while negative and quite large, the effect for Asia was only significant at the 14% level).

For Europe, we have the advantage of observing a separate, more detailed matrix of F_{ST} genetic distance.¹⁵ The results were strong in this subsample: despite the paucity of observations (only 291 country pairs), the effect of genetic distance remained negative and significant at the 1% level, and its standardized magnitude was much larger than in the worldwide sample. European countries are geographically very connected, either by land or sea, so genetic distance is unlikely to capture geographic impediments to conflict. Moreover, genetic distance in Europe results from much more recent population divisions. To be able to identify a large positive effect of relatedness on conflict

¹⁵Estimates using the European matrix, where there are 26 distinct genetic groups, are based on more precise measures compared to the worldwide sample, as detailed in Spolaore and Wacziarg (2009). More extensive estimation results focusing on Europe are available in the online appendix.

propensities even among populations that are closely related historically reinforces the robustness of our main result. Nonetheless, the worldwide results were not driven by Europe: in a sample including only pairs of non-European countries, the negative effect of genetic distance remained large and statistically significant.

We next examined whether relatedness affects conflict differently across time. To do so, we defined dummy variables for whether a country pair was ever in conflict during a specific subperiod.¹⁶ Results are presented in the online appendix. We found that the estimated effect of genetic distance is remarkably robust across time periods: it remains negative, large and significant whether considering the pre- or post-1900 periods, suggesting that our findings are not driven by the wars of the 20th century, in particular the two World Wars. Focusing on the 20th century, the effect is unchanged for the post-1946 period, compared to the 1816 – 2001 baseline. In other words, our finding is not simply an artifact of the Second World War, which pitted a lot of European populations against each other. Our finding holds significantly even after the end of the Cold War, despite the relatively small number of pairs involved in conflicts during this period (only 218).

3.4 Linguistic, Religious, and Cultural Distances

While genetic distance is a precise and continuous measure of the degree of relatedness between populations and countries, other measures exist. Linguistic relatedness is associated with genetic relatedness because, like genes, languages are transmitted intergenerationally: populations speaking similar languages are likely to be more related than linguistically distinct populations (Cavalli-Sforza et al., 1994). On the other hand, there are many reasons why genetic and linguistic distance are imperfectly correlated. Rates of genetic and linguistic mutations may differ. Populations of a certain genetic make-up may adopt a foreign language as the result of foreign rule, as happened when the Magyar rulers imposed their language in Hungary. Other salient examples include countries colonized by European powers, adopting their language (English, French, Portuguese or Spanish), while maintaining distinct populations genetically.

Religious beliefs, also transmitted intergenerationally, are a type of human traits that can affect conflict. In fact, the existing literature on interstate conflict has examined linguistic and religious ties in an effort to test primordialist theories of conflict (Richardson, 1960, Henderson, 1997). Thus,

¹⁶These subperiods, and the corresponding number of pairs that were involved in conflict during those subperiods, were: 1816-1900 (106 pairs in conflict), 1901-2001 (721 pairs in conflict), 1946-2001 (536 pairs in conflict), 1919-1989 (585 pairs in conflict), 1990-2001 (218 pairs in conflict).

it is important to evaluate whether these variables trump genetic distance, and more generally how their inclusion affects our main coefficient of interest. In what follows, we evaluate whether the effect of genetic distance is reduced or eliminated when controlling for linguistic and religious distance, and whether these variables have an independent effect on the incidence of interstate conflict.

To capture linguistic distance, we use data from Fearon (2003), based on linguistic trees from Ethnologue to compute the number of common linguistic nodes between languages in the world. To measure religious distance we follow an approach based on religious trees, using a nomenclature of world religions obtained from Meham, Fearon and Laitin (2006).¹⁷ Table 7 presents estimates of the effect of genetic distance on the propensity for interstate conflict when linguistic and religious distances are included. Since the use of these variables constrains the sample (a loss of some 3,154 observations, or almost 24% of the sample), we start in column (1) with the baseline estimates for this new sample. They are in line with those reported above. When adding linguistic distance and religious distance either alone or together (columns 2-4), interesting results emerge. First, the coefficient on genetic distance is barely affected. Second, linguistic distance exerts a null effect when controlling for genetic distance. Third, religious distance is negatively associated with conflict, and this effect is statistically significant even when including linguistic distance along with religious distance.¹⁸ The latter finding is consistent with the view that religion is one of the vertically transmitted traits that make populations more or less related to each other, and its effect on conflict goes in the same direction as that of genetic distance, a broader measure of relatedness.

Finally, we also included in the baseline specification the measures of cultural distance described in Section 2.2 (these regression results are available in the online appendix). The summary index based on responses to all 98 WVS questions entered with a negative sign, as expected, and its inclusion reduced the effect of genetic distance. When including the cultural distance indices specific to each WVS question category all together, the effect of genetic distance was reduced further, again consistent with our hypothesis that genetic distance captures a broad array of human traits

¹⁷These measures are further described in Spolaore and Wacziarg (2009). Pairwise correlations between measures of genetic, linguistic and religious distances are positive, as expected, but not very large. For instance, the correlation between F_{ST} genetic distance and linguistic distance is 0.201. Religious distance bears a correlation of 0.449 with linguistic distance, and 0.172 with genetic distance.

¹⁸This result contrasts with that in Henderson (1997), who found evidence that religious *similarity* was negatively related to conflict. The difference may stem from our use of a different (and more fine-grained) measure of religious distance, our much bigger sample, as well as our inclusion of a much broader set of controls (Henderson only controlled for contiguity).

transmitted from generation to generation.

3.5 Extensions

We briefly discuss three extensions, for which results are available in the online appendix. First, we investigated nonlinear effects of genetic distance, and interaction effects with geographic distance, finding scant evidence of such effects. Second, we sought to evaluate the effect of genetic distance on the intensity of conflict, finding that most of the effect stems from the extensive margin: conditional on conflict, genetic distance does not predict its intensity. Third, we conducted a panel analysis of the determinants of a yearly indicator of bilateral conflict. Since genetic distance does not vary by time, the only benefit of this analysis is to explore the robustness of our result to controlling for time-varying factors such as international trade, democracy levels, and income differences. We found that genetic distance continues to be a significantly negative determinant of conflict after controlling for these time-varying factors. We also found that the main claims of liberal peace theory, namely that democracies tend not to fight with each other and that bilateral trade has a pacifying effect, continue to hold even after controlling for genetic distance.

4 Evidence on Mechanisms

4.1 Natural Endowments, Relatedness and Conflict

If wars are about specific rival goods, such as oil, an indicator variable taking a value of 1 if one of the countries in the pair possesses that rival good and 0 otherwise should have a positive effect on the probability of conflict. Moreover, if the mechanism we highlight is operative, the interaction term between the presence of oil and genetic distance should take a negative sign, because countries with more closely related populations would share more similar preferences and therefore should be more likely to fight over that rival good. In sum, when running:

$$C_{ij} = \beta_1 X_{ij} + \beta_2 FST_{ij}^W + \beta_3 RIV_{ij} + \beta_4 FST_{ij}^W \times RIV_{ij} + \eta_{ij} \quad (3)$$

where $RIV_{ij} = 1$ if either i or j (or both) has a lot of a specific rival good (e.g., oil), we should expect $\beta_3 \geq 0$ and $\beta_4 \leq 0$.¹⁹

¹⁹The inequalities are not strict. β_3 and β_4 could be zero if this specific rival good does not matter for conflict in our sample: our framework does not predict that countries actually go to war about each possible rival good.

To conduct this test we examine three rival goods over which countries may fight. The first is oil. Oil became a crucial source of energy since the widespread diffusion of the internal combustion engine, so for this measure we confine attention to conflicts after 1945. We define a dummy variable, $RIV_{ij} = OIL_{ij}$, equal to 1 if either i , j or both were major producers of oil (defined as producing an average of more than 150,000 barrels per day of crude oil) in the 1980s, 1990s and 2000s (the data is from Wacziarg, 2012). The second variable is the presence of fertile land, expected to have been a desirable feature of countries in the period when agriculture represented a large portion of total output. We define a dummy variable, $RIV_{ij} = LAND_{ij}$, equal to 1 if in either country j or i (or both) at least 40% of land area is fertile soil (the data is from Nunn and Puga, 2012). Finally we consider the share of land located in a temperate climate. The dummy variable $RIV_{ij} = TEMP_{ij}$ is equal to 1 if either country in a pair (or both) has more than 60% of its land area located in a temperate Köppen-Geiger climatic zone (defined, as zones Cf, Cs, Df and DW). For the latter two measures of rival goods, we confine attention to explaining conflicts that occurred prior to 1900, when soil and climate were more likely to matter.

The results, presented in Table 6, are consistent with the mechanism we highlight to explain the negative relationship between genetic distance and the probability of conflict. To allow comparisons, for each of our three measures of rival goods, we include both a baseline specification omitting the dummy for the rival good and its interaction with genetic distance, and one that includes them, using the same sample. The first two columns refer to the oil sample. With the oil dummy and the interaction term included (column 2), the pattern is precisely consistent with our mechanism: the oil dummy itself is a significant determinant of conflict, while its interaction with genetic distance is negative. The standardized magnitude of the total effect of genetic distance, when the oil dummy is equal to 1, is 28.05% as opposed to 19.69% in the baseline, but genetic distance on its own remains significant even with the oil terms added, suggesting that genetic distance does not capture only the effect of oil conflicts. We find similar results for the temperate climate dummy and the fertile soil dummy between 1816 and 1900: the presence of these attributes raises the probability of interstate conflict and their interaction with genetic distance is negative, supporting our interpretation that similarity in preferences raises the probability of conflict when there is a desirable feature of territories to fight over.

4.2 Territorial Changes, Independence, and Decolonization

Territory is the quintessential example of a rival good at the center of international conflict. We expect that more similar populations have incentives to fight more intensively over territories with more similar physical and human characteristics. Similarity in preferences and behaviors among related populations also creates stronger incentives for rulers to conquer territories inhabited by populations that are closer to those they already rule. When going to war, rulers often think not only about "winning the war" but also about "winning the peace": after a war, governments will find it easier to rule over more closely related groups, with closer views about types of government and policies, than over more dissimilar groups.²⁰

If the mechanism we highlight is operative, peaceful territorial changes should take place between states that are less closely related, while violent territorial change should occur between more closely related states. In particular, we expect more peaceful transitions to independence for populations under the control of more distantly related states, while violent conflict in the breakup of countries and empires should occur more often when the secessionist periphery and the central government are more closely related. Finally, we expect that these effects played a key role in the process of decolonization, when colonial powers lost territories (dependencies) to new states formed by populations that gained control of their own homeland. To test these implications of our theory, we use Version 4.01 of the dataset on territorial changes included as part of the Correlates of War (Tir et al. 1998).

4.2.1 Violent and Non-Violent Territorial Change

We first consider all instances of territorial change between 1816 and 2008 in the COW territorial change dataset, provided that a territory is transferred to a sovereign state 1 (the gainer) from a sovereign state 2 (the loser), and that both states are in our genetic-distance dataset. There are 545 such territorial transfers in the sample. Most of these transfers (409 observations, 75% of the total) took place peacefully, while 136 (25% of the total) were the outcome of violent conflict. As shown in Table 7, the mean genetic distance between the gainer and the loser in peaceful territorial changes is 0.068, while the genetic distance between gainer and loser in violent transfers is significantly lower, and equal to 0.050. This finding is fully consistent with our interpretation of interstate conflict as emerging from territorial disputes between more similar rulers and populations.

²⁰We thank a referee for this observation and terminology.

4.2.2 Territorial Change and Independence

An important instance of territorial change is when the gainer is a new state, which therefore becomes independent from the loser. Gainers and losers in the 20th century include Ireland and the UK in 1922, Algeria and France in 1962, Eritrea and Ethiopia in 1993, and many more. By definition, conflicts over independence took place not between two existing independent states, but between state rulers (such as the UK or France) and populations that had not yet formed their own independent country (such as the Irish or the Algerians). Since these cases are not part of our dataset of interstate conflict, they provide an out-of-sample test of our mechanism.

Out of the 545 territorial changes between 1816 and 2008 described above, the COW data set identifies 156 instances that resulted in independence for the gainer. These events provide a direct test of our hypothesis that a central government would be much more reluctant to concede independence to genetically closer populations, while it would be less unwilling to let more distant populations form their own state. Out of 156 territorial changes resulting in independence, 124 cases (79.5% of the total) took place peacefully, while only 32 (20.5%) were the outcome of violent conflict. Consistent with our hypothesis, the average genetic distance between newly independent state and loser state is much higher in the non-violent cases (0.102) than in the violent cases (0.053), as shown in Table 7. These data are instructive for two reasons. First, they document that most states in the past two centuries gained independence peacefully rather than violently. Second, the relatively rarer instances of violent conflicts over independence happened when secessionist populations were much closer genetically - and therefore historically and culturally - to the central government that resisted their demand, consistent with the mechanisms we have emphasized.

4.2.3 Conflicts over Homeland Territories and Colonial Dependencies

A useful distinction in the COW data set is whether territorial change involved colonial dependencies or territories that were part of a central state's homeland. For example, when France lost Alsace to Germany in 1871, and gained it back in 1919, the Alsatian territory was considered part of the homeland of both countries. In contrast, when Pondicherry and other territories were transferred from France to India in 1954, India gained homeland territories while France lost colonial dependencies. Therefore, this classification allows us to study more closely the process of decolonization.

Out of 545 territorial changes between 1816 and 2008, 196 instances took place between a state that gained a homeland territory and a state that lost a colonial dependency. The overwhelming

majority of territorial changes associated with decolonization took place peacefully. We observe 162 peaceful changes out of 196 (82.7% of the total), while only 34 instances (17.3% of the total) occurred as a result of violent conflict. Why was decolonization so peaceful overall? A key mechanism at work to ensure a mostly peaceful decolonization was indeed the high historical and cultural distance between former colonial powers and their dependencies, measured by genetic distance. This implication can be directly examined in our dataset: the mean genetic distance between homeland gainer and colonial loser when the territorial change took place without violent conflict was 0.098. In contrast, in the fewer cases of violent conflict, the mean genetic distance was only 0.058, as shown in Table 7. This evidence supports our hypothesis that colonial rulers were less likely to go to war over territories and populations that were dissimilar from themselves. The relation also holds when we focus on the 138 instances of territorial change that led to independence in former colonies (Table 7). In the 109 peaceful cases of independence for former colonies, which are 79% of the total, the mean genetic distance was 0.110, while in the remaining 29 cases (21% of the total) when decolonization was violent, average genetic distance between colonizers and former colonies was 0.057. These findings are in line with the general results for territorial changes associated with independence discussed above.

A similar, although less dramatic, pattern holds for territorial changes involving the homeland of both gainer and loser. In those cases, we observe 129 territorial changes without violent conflict and 78 territorial changes with violent conflict. The mean genetic distance is 0.050 in the absence of violent conflict and 0.039 in the presence of violent conflict (Table 7), consistent with the hypothesis that closer relatedness spurs conflict over rival goods. For instance, the recent conflict between Russia and Ukraine over Crimea and the Eastern region of Ukraine (not included in our sample, which stops in 2008) is consistent with our prediction that states with closely related populations are more likely to engage in conflict over control of rival goods and populations. A discussion of the cultural and religious importance of Crimea in both Russian and Ukrainian history is provided by Turchin (2014).

5 Conclusion

We examined the empirical relationship between the occurrence of interstate conflicts and the degree of relatedness between countries, and found that populations that are closer in terms of genetic, linguistic and religious distances are more prone to engage in militarized conflicts across national borders. This effect is large in magnitude and robust to controlling for several measures of geographic

distance, income differences, and other factors affecting conflict.

We provided an economic interpretation of these results. The central insight is that populations that are more closely related tend to share more similar preferences over resources and territories, and are therefore more likely to enter into conflicts. In particular, populations that share more similar preferences over rival goods tend to care about the same natural resources. For example, we showed that the presence of oil is associated with a higher likelihood of conflict between countries after 1945, and that the effect is larger for countries that are genetically closer. Similar effects hold for fertile soil and temperate climate before 1900.

Territory is the ultimate rival good at the center of international conflict. More closely related groups can be expected to fight more intensively over territories with more similar physical and human characteristics, while similarity in preferences and views among related populations creates stronger incentives for rulers to conquer territories inhabited by populations that are closer to those they already rule. Data on territorial changes across countries provide evidence consistent with our interpretation. Most territorial changes between 1816 and 2008 took place peacefully, while the rarer instances of violent changes, including violent processes of independence and decolonization, occurred when the ruling country and the break-away region were inhabited by more closely related populations.

Our paper is about international conflict. In principle, the effects of relatedness on conflict can be different when considering civil or ethnic conflict within a country. Less similar groups within a country can be expected to fight less over specific rival goods. However, they may also fight more over common goods and policies that all must share in spite of their different preferences and traits. Insofar as civil conflict is about the control of a common central government by different groups, we should observe an opposite relation between relatedness and conflict within a given country - that is, less relatedness leading to more civil conflict. In contrast, civil conflict over rival resources within a country would tend to take place among very closely related groups, sharing similar preferences.²¹

Our findings provide evidence against the primordialist view that cultural dissimilarity between

²¹This theoretical ambiguity is reflected in the ongoing debate on the role of ethnic divisions in fostering civil conflict, including contributions by Fearon and Laitin (2003), Montalvo and Reynal-Querol (2005), Desmet, Ortuño-Ortín and Wacziarg (2012, 2015) among many others. In a recent empirical study of ethnicity and intrastate conflict, Esteban, Mayoral and Ray (2012) find that when civil conflict is mostly over public goods, ethnolinguistic polarization leads to more conflict. This finding is consistent with the hypothesis that less closely related groups are more likely to fight over the control of a common government and policies.

countries should breed war, plunder and a clash of civilizations. On the contrary, relatedness is positively associated with interstate conflict, and closer sibling states are more likely to fight with each other than with their more distant cousins. From a long-term historical perspective, issues of war and peace across nations are family matters.

References

- Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003), "Fractionalization," *Journal of Economic Growth*, 8: 55-194.
- Alesina, A. and E. Spolaore (2006), "Conflict, Defense Spending, and the Number of Nations," *European Economic Review* 50(1), 91-120.
- Arbatli, C. E., Q. Ashraf and O. Galor (2015), "The Nature of Civil Conflict," NBER Working Paper No. 21079, April.
- Ashraf, Q. and O. Galor (2013), "The 'Out-of-Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103(1): 1-46.
- Bellwood, Peter (2013), *First Migrants: Ancient Migration in Global Perspective*, Malden (MA): Wiley-Blackwell.
- Bisin, A. and T. Verdier (2001), "The Economics of Cultural Transmission and the Evolution of Preferences," *Journal of Economic Theory*, 97: 98-319.
- Boulding, K. E. (1962), *Conflict and Defense: A General Theory*, New York: Harper.
- Bowles, S. (2009), "Did Warfare Among Ancestral Hunter-gatherers Affect the Evolution of Human Social Behavior?" *Science*, June, 324 (5932): 1293-1298.
- Bremer, S. A. (1992), "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War 1816-1965," *Journal of Conflict Resolution*, 36 (2): 309-341.
- Bremer, S. A. (2000), "Who Fights Whom, When, Where, and Why?" in J. A. Vasquez, ed. (2000), *What Do We Know About War?* Lanham: Rowman & Littlefield.
- Caselli, F., M. Morelli and D. Rohner (2015), "The Geography of Inter-State Resource Wars," *Quarterly Journal of Economics*, 130 (1): 267-315.
- Cavalli-Sforza, L. L., P. Menozzi and A. Piazza (1994), *The History and Geography of Human Genes*, Princeton: Princeton University Press.
- Desmet, K., M. Le Breton, I. Ortuño-Ortín and S. Weber (2011), "The Stability and Breakup of Nations: a Quantitative Analysis," *Journal of Economic Growth*, 2011, 16, pp. 183-213.

- Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012), "The Political Economy of Ethnolinguistic Cleavages," *Journal of Development Economics*, 97 (1): 322-32.
- Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2015), "Culture, Identity and Diversity," *NBER Working Paper #20989*, February.
- Dyen, I., J. B. Kruskal and P. Black (1992), "An Indo-European Classification: A Lexicostatistical Experiment," *Transactions of the American Philosophical Society*, 82: 1-132.
- Esteban, J., L. Mayoral and D. Ray (2012), "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, 102(4): 1310-1342.
- Faten G., G. Palmer, and S. A. Bremer (2004), "The MID3 Data Set, 1993-2001: Procedures, Coding Rules, and Description," *Conflict Management and Peace Science*, 21: 133-154.
- Fearon, J. (2003), "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 8: 195-222.
- Fearon, J. and D. Laitin (2003), "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 97(1): 75-90.
- Garfinkel, M. R. and S. Skaperdas (2007), "Economics of Conflict: An Overview," in K. Hartley and T. Sandler (eds) *Handbook of Defense Economics*, Volume 2, North Holland, Amsterdam, 649-709.
- Gleditsch, N. P. and J. D. Singer (1975), "Distance and International War, 1816-1965," in *Proceedings of the International Peace Research Association (IPRA) Fifth General Conference (Oslo: IPRA)*: 481-506.
- Henderson, E. A. (1997), "Culture or Contiguity. Ethnic Conflict, the Similarities of States, and the Onset of War, 1820-1989," *Journal of Conflict Resolution*, 41, 5, 649-668.
- Heston, A., R. Summers, and B. Aten (2002), "Penn World Table Version 6.1," *Center for International Comparisons at the University of Pennsylvania (CICUP)*.
- Hirshleifer, J. (1998), "The Bioeconomic Causes of War," *Managerial and Decision Economics*, 19(7/8): 457-466.
- Huntington, S. (1993), "The Clash of Civilizations," *Foreign Affairs* 72(3): 22-49.
- Jones, D. M., S. A. Bremer, and J. D. Singer (1996), "Military Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Patterns," *Conflict Management and Peace Science*, 15: 163-213.
- Martin, P., T. Mayer and M. Thoenig (2008) "Make Trade Not War?" *Review of Economic Studies*, 75: 865-900.
- Mecham, R. Q., J. Fearon and D. Laitin (2006), "Religious Classification and Data on Shares of

- Major World Religions," *unpublished*, Stanford University.
- Montalvo, J. G. and M. Reynal-Querol (2005), "Ethnic Polarization, Potential Conflict and Civil War," *American Economic Review*, 95(3): 796-816.
- Nunn, N. and D. Puga (2012), "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 94(1), pp. 20-36.
- Oster, E. (2014), "Unobservable Selection and Coefficient Stability: Theory and Evidence," *working paper*, University of Chicago, January.
- Özak, Ö. (2013), "Distance to the Technological Frontier and Economic Development," *working paper*, Southern Methodist University.
- Putterman, L. and D. Weil (2010), "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, vol. 125, no. 4, pp. 1627-1682.
- Richardson, L. (1960), *Arms and Insecurity*, Pacific Grove: Boxwood.
- Richerson P. J. and R. Boyd (2004), *Not By Genes Alone: How Culture Transformed Human Evolution*, Chicago: University of Chicago Press.
- Schelling, T. (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
- Singer, J. D. (1972), *The Wages of War, 1816-1965: A Statistical Handbook*, New York: Wiley.
- Singer, J. D., S. Bremer and J. Stuckey (1972), "Capability Distribution, Uncertainty, and Major Power War, 1820-1965," in Bruce Russett, ed, *Peace, War, and Numbers*, Beverly Hills: Sage.
- Spolaore E. and R. Wacziarg (2009), "The Diffusion of Development," *Quarterly Journal of Economics*, 124(2): 469-529.
- Spolaore E. and R. Wacziarg (2013a), "War and Relatedness," *Working Paper, UCLA, February*.
- Spolaore E. and R. Wacziarg (2013b), "How Deep are the Roots of Economic Development?," *Journal of Economic Literature*, 51(2): 325-69.
- Spolaore E. and R. Wacziarg (2015), "Ancestry, Language and Culture," forthcoming in Victor Ginsburgh and Shlomo Weber, eds, *Palgrave Handbook of Economics and Language*, London: Palgrave Macmillan.
- Sumner, W. G. (1906), *Folkways*, New York: Ginn.
- Tir, J., P. Schafer, P. F. Diehl and G. Goertz (1998), "Territorial Changes, 1816-1996: Procedures and Data," *Conflict Management and Peace Science* 16, 89-97.
- Turchin, P. (2014), "Russia's Sacred Land," *Aeon*, April 3.
- Vasquez, J. A. (1995), "Why Neighbors Fight? Proximity, Interaction or Territoriality," *Journal of Peace Research*, 32(93): 277-293.

Vasquez, J. A. and M. T. Henehan (2001), "Territorial Disputes and the Probability of War, 1816-1992," *Journal of Peace Research*, 38(2): 123-138.

Wacziarg, R. (2012), "The First Law of Petropolitics," *Economica*, vol. 79, no. 316, pp. 641-657.

Wright, Q. (1942), *A Study of War*, Chicago: University of Chicago Press.

Table 1 – Summary statistics and correlations for major variables

Panel A – Summary Statistics

Variable	# Obs.	Mean	Std. Dev.	Min	Max
Conflict (%)	13,175	0.056	0.231	0	1
War (%)	13,175	0.021	0.143	0	1
Fst genetic distance, weighted	13,175	0.111	0.068	0	0.355
Log geodesic distance	13,175	8.700	0.787	2.349	9.899
Dummy for contiguity	13,175	0.019	0.136	0	1
Religious Distance Index, weighted	10,155	0.846	0.149	0.089	1
Linguistic Distance Index, weighted	10,021	0.968	0.107	0	1

Panel B – Pairwise Correlations

	Conflict (%)	War (%)	FST genetic distance	Log geodesic distance	Contiguity	Religious distance
War (%)	0.597* (13,175)	1 (13,175)				
Fst genetic distance, weighted	-0.169* (13,175)	-0.107* (13,175)	1 (13,175)			
Log geodesic distance	-0.217* (13,175)	-0.105* (13,175)	0.434* (13,175)	1 (13,175)		
Dummy for contiguity	0.337* (13,175)	0.164* (13,175)	-0.146* (13,175)	-0.362* (13,175)	1 (13,175)	
Religious Distance Index, weighted	-0.132* (10,155)	-0.052* (10,155)	0.168* (10,155)	0.211* (10,155)	-0.140* (10,155)	1 (10,155)
Linguistic Distance Index, weighted	-0.140* (10,021)	-0.073* (10,021)	0.201* (10,021)	0.240* (10,021)	-0.194* (10,021)	0.449* (10,021)

(# of observations in parentheses; * denotes significance at the 5% level)

Table 2 – Distribution of War and Conflict, by Quartile of Genetic Distance

Conditioning statement:	Bottom decile of genetic distance	0-25th percentile of genetic distance	25-50th percentile of genetic distance	50-75th percentile of genetic distance	75-100th percentile of genetic distance*	Total
Hostility level = 5 (War)						
None	76 27.64%	148 53.82%	70 25.45%	47 17.09%	10 3.64%	275 100%
Common sea / ocean = 0	49 22.79%	107 49.77%	56 26.05%	42 19.53%	10 4.65%	215 100%
Contiguity = 0	52 22.81%	117 51.32%	55 24.12%	46 20.18%	10 4.39%	228 100%
Distance > 1000 km	54 23.28%	119 51.29%	56 24.14%	47 20.26%	10 4.31%	232 100%
Hostility Level > 3 (Conflict)						
None	188 25.27%	400 53.76%	195 26.21%	103 13.84%	46 6.18%	744 100.00%
Common sea / ocean = 0	123 22.65%	283 52.12%	138 25.41%	81 14.92%	41 7.55%	543 100.00%
Contiguity = 0	124 20.98%	297 50.25%	153 25.89%	96 16.24%	45 7.61%	591 100.00%
Distance > 1000 km	119 19.41%	301 49.10%	165 26.92%	101 16.48%	46 7.50%	613 100.00%

Based on an underlying sample of 13,175 country pairs.

* 7 of the 10 cases in rows 3-6 involve South Africa as a combatant.

Table 3 - Cross sectional regressions, probit or IV probit estimator (1816-2001 sample)
(Dependent variable: dummy for whether a country pair was ever involved in a conflict or war between 1816 and 2001)

	(1)	(2)	(3)	(4)	(5)	(6)
	Conflict, univariate specification	Conflict, baseline specification	Conflict, baseline specification IV	Conflict, non-contiguous pairs only	War, baseline specification	War, baseline specification IV
Fst genetic distance, weighted	-57.3760** (-17.800)	-19.8786** (-9.317)	-30.6802** (-8.843)	-18.5357** (-9.379)	-6.3389** (-7.478)	-8.6043** (-5.746)
Log geodesic distance		-1.6281** (-5.567)	-1.0182** (-3.090)	-1.4809** (-5.065)	-0.2929* (-2.505)	-0.1728 (-1.349)
Log absolute difference in longitudes		0.1424 (0.731)	-0.0677 (-0.336)	0.1629 (0.842)	-0.0197 (-0.254)	-0.0629 (-0.787)
Log absolute difference in latitudes		-0.1130 (-0.887)	-0.1312 (-1.002)	-0.0729 (-0.614)	-0.1314** (-2.612)	-0.1366** (-2.660)
1 for contiguity		15.4610** (10.095)	16.2256** (5.465)	-	0.8262** (2.701)	0.9060 (1.856)
Number of landlocked countries in the pair		-2.6247** (-9.471)	-2.6311** (-9.566)	-2.4127** (-8.927)	-0.6406** (-5.531)	-0.6500** (-5.635)
Number of island countries in the pair		0.8212** (2.923)	0.8762** (3.005)	0.6967** (2.755)	0.4118** (3.828)	0.4439** (3.711)
1 if pair shares at least one sea or ocean		1.9440** (4.909)	1.9435** (3.799)	1.9330** (5.181)	-0.0154 (-0.128)	-0.0199 (-0.161)
Log product of land areas in square km		0.8940** (18.992)	0.9045** (17.145)	0.7960** (18.528)	0.3132** (17.452)	0.3201** (9.755)
1 for pairs ever in colonial relationship		7.3215** (5.094)	7.6147** (3.175)	8.6303** (6.004)	0.9013* (2.099)	0.9754 (1.463)
1 if countries were or are the same country		1.9512 (1.846)	2.2217 (1.541)	1.6352 (1.229)	1.0952* (2.424)	1.1373 (1.564)
# of observations	13,175	13,175	13,175	12,928	13,175	13,175
Pseudo-R ²	0.075	0.275	-	0.202	0.236	-
Standardized effect (%)	-68.81	-23.84	-36.79	-27.34	-20.57	-27.92

Robust t statistics in parentheses; * significant at 5%; ** significant at 1%.

The standardized magnitude refers to the effect of a one-standard deviation increase in genetic distance as a percentage of the mean probability of conflict/war for the sample used in each regression.

Probit marginal effects are reported in all columns. For dummy variables, marginal effects are for discrete changes from 0 to 1. All marginal effects were multiplied by 100 for readability.

**Table 4 – Baseline Specification with Additional Geographic Controls, probit estimator (1816-2001 sample)
(Dependent variable: dummy for whether a country pair was ever involved in a conflict between 1816 and 2001)**

	(1)	(2)	(3)	(4)	(5)
	Nunn-Puga controls	Climatic conditions	Elevation difference	Ozak distance	All together
Fst genetic distance, weighted	-16.7215** (-9.141)	-23.4126** (-7.968)	-27.8133** (-8.977)	-25.3596** (-8.751)	-20.3039** (-7.453)
Abs. difference in ruggedness index, pop weighted	0.5463* (2.486)				0.7732* (2.286)
Abs. difference in % tropical area	-0.0079** (-2.729)				-0.0160** (-3.526)
Abs. difference in % fertile soil	-0.0108* (-2.046)				-0.0088 (-1.115)
Abs. difference in % desert area	-0.0806** (-7.753)				-0.1049** (-7.101)
Abs. difference in avg. distance to coast	-1.2599** (-4.088)				-2.1732** (-4.939)
Abs. difference in % within 100 km. of ice-free coast	0.0140** (3.469)				0.0287** (4.805)
Measure of climatic difference of land areas, by 12 KG zones		0.0469 (1.136)			0.1122** (2.758)
Absolute value of difference in average elevation, 1000s of ft			0.2333 (0.711)		-0.1838 (-0.588)
HMisea cost (weeks), avg. of both directions, Ozak				-0.0499* (-1.980)	-0.0742** (-3.035)
# of pairs	13,033	10,216	10,492	10,230	9,055
Pseudo-R ²	0.300	0.255	0.253	0.276	0.302
Standardized effect (%)	-20.00	-25.34	-28.02	-26.29	-21.19

Robust t statistics in parentheses; * significant at 5%; ** significant at 1%.

Probit marginal effects reported in all columns. For dummy variables, marginal effects are for discrete changes from 0 to 1. All marginal effects were multiplied by 100 for readability. The standardized magnitude is the effect of a one standard deviation increase in genetic distance as a percentage of the mean probability of conflict.

Controls: In addition to reported coefficients, every column includes controls for: Log geodesic distance, log absolute difference in longitudes, Log absolute difference in latitudes, dummy for contiguity, number of landlocked countries in the pair, number of island countries in the pair, dummy=1 if pair shares at least one sea or ocean, log product of land areas in square km, dummy for pairs ever in colonial relationship, dummy=1 if countries were or are the same country.

Table 5 – Adding other measures of historical distance
(Dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(4)
	Baseline specification	Add linguistic distance	Add religious distance	Add religious and linguistic distances
Fst genetic distance, weighted	-29.3281 (8.872)**	-29.1266 (8.792)**	-27.1691 (8.369)**	-27.4118 (8.484)**
Log geodesic distance	-2.4924 (5.374)**	-2.4971 (5.379)**	-2.4498 (5.315)**	-2.4268 (5.291)**
1 for contiguity	22.5037 (10.375)**	22.3377 (10.308)**	21.4007 (10.161)**	21.7116 (10.155)**
Linguistic Distance Index, weighted	-	-0.8099 (0.659)	-	2.3819 (1.778)
Religious Distance Index, weighted	-	-	-5.1999 (5.013)**	-5.9958 (5.281)**
Pseudo-R ²	0.250	0.250	0.255	0.255
Standardized effect (%)	-28.050	-27.857	-25.985	-26.217

Robust t statistics in parentheses; * significant at 5%; ** significant at 1%.

The standardized magnitude is the effect of a one standard deviation increase in genetic distance as a percentage of the mean probability of conflict. The table reports marginal effects from probit estimates. For dummy variables, marginal effects are for discrete changes from 0 to 1. All coefficients were multiplied by 100 for readability.

10,021 observations used in all columns.

Controls: In addition to reported coefficients, all regressions include controls for log absolute difference in longitudes, log absolute difference in latitudes, number of landlocked countries in the pair, number of island countries in the pair, dummy for pair shares at least one sea or ocean, log product of land areas in square km, dummy for pairs ever in colonial relationship, dummy for countries were or are the same country.

Table 6: Interactions with Oil, Temperate Climate, and Fertile Soil
 (Dependent variable: dummy for whether a country pair was ever in conflict in the period specified in row 3)

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline (oil sample)	Oil	Baseline (temp. climate sample)	Temperate climate	Baseline (fertile soil sample)	Fertile Soil
	1945-2001	1945-2001	1816-1900	1816-1900	1816-1900	1816-1900
Fst genetic distance, weighted	-11.8279** (-6.933)	-7.1885** (-3.184)	-2.0078** (-5.470)	-0.1500 (-0.549)	-1.0450** (-5.396)	-0.0079 (-0.054)
Log geodesic distance	-1.0813** (-5.185)	-1.1553** (-5.361)	-0.0853 (-1.455)	0.0136 (0.331)	-0.0750* (-2.571)	-0.0339 (-1.697)
Interaction of oil producer dummy and gen. dist.		-9.6647** (-3.124)				
Dummy for at least one country in the pair being a major oil producer		1.3988** (3.833)				
Interaction of temperate climate and gen. dist.				-1.2588** (-2.886)		
Dummy for one or more country in the pair with >60% land in temperate zone				0.6627** (8.386)		
Interaction of fertile soil dummy and gen. dist.						-0.8456** (-4.069)
Dummy for one or more country in the pair with >40% fertile soil						0.1381** (7.510)
# of observations	13,175	13,175	10,216	10,216	13,033	13,033
Pseudo-R ²	0.280	0.284	0.261	0.313	0.289	0.322
Standardized effect ^a	-19.690	-28.050	-15.470	-10.850	-8.632	-7.050

Robust z-statistics in parentheses; ** p<0.01, * p<0.05.

The standardized magnitude refers to the effect of a one-standard deviation increase in genetic distance as a percentage of the mean probability of conflict/war for the sample used in each regression.

^a: with interaction effects the standardized magnitude is the *total* standardized effect of genetic distance when the endowment dummy is equal to 1.

Probit marginal effects are reported in all columns. For dummy variables, marginal effects are for discrete changes from 0 to 1. All marginal effects were multiplied by 100 for readability.

Additional included controls (estimates not reported): log absolute difference in longitudes, log absolute difference in latitudes, dummy=1 for contiguity, number of landlocked countries in the pair, number of island countries in the pair, dummy=1 if pair shares at least one sea or ocean, log product of land areas in square km, dummy=1 for pairs ever in colonial relationship, dummy=1 if countries were or are the same country.

Table 7 – Analysis of the Territorial Changes Dataset

Conflict status	# of observations	Mean Genetic Distance
All territorial changes		
conflict = 0	409	0.0676
conflict = 1	136	0.0500
All territorial changes resulting in independence for the gainer		
conflict = 0	124	0.1023
conflict = 1	32	0.0526
All territorial changes resulting in independence for the gainer, the loser losing a colonial dependency and the gainer gaining a homeland		
conflict = 0	109	0.1097
conflict = 1	29	0.0567
All territorial changes, the loser losing a colonial dependency and the gainer gaining a homeland		
conflict = 0	162	0.0977
conflict = 1	34	0.0584
All territorial changes, both loser and gainer gaining a homeland		
conflict = 0	129	0.0498
conflict = 1	78	0.0388

Source: Correlates of War project, Territorial Changes Dataset, version 4.01 (Tir et al., 1998)