THE UNIVERSITY OF WARWICK

June 2013 No.140

War and Relatedness

Enrico Spolaore Tufts University, NBER, CESIfo and CAGE
Romain Wacziarg UCLA, NBER and CEPR

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE
Competitive Advantage in the Global Economy

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

War and Relatedness*

Enrico Spolaore

Tufts University, NBER, CESIfo and CAGE

Romain Wacziarg

UCLA, NBER and CEPR

This version: January 2013

Abstract

We examine the empirical relationship between the occurrence of inter-state conflicts and the degree of relatedness between countries, measured by genetic distance. We find that populations that are genetically closer 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, including measures of trade and democracy. These findings are consistent with a framework in which conflict over rival and excludable goods (such as territory and resources) is more likely among populations that share more similar preferences, and inherit such preferences with variation from their ancestors.

*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 Don Cox, Klaus Desmet, Rajeev Dehejia, Quy-Toan Do, Steven Durlauf, Jim Fearon, Jeff Frieden, Oded Galor, Michelle Garfinkel, Tim Gloseclose, Yannis Ioannides, Michael Klein, Glenn Loury, Philippe Martin,Deborah Menegotto, Massimo Morelli, Daniele Paserman, Ken Shepsle, Vickie Sullivan and seminar participants at Stanford University, UCLA, Dartmouth College, Claremont McKenna College, Brown University, and numerous conferences for helpful comments. We gratefully acknowledge financial support from Stanford University's Presidential Fund for Innovation in International Studies and from UCLA's Center for International Business Education and Research. This is a substantially revised version of the paper previously circulated under the same title.

1 Introduction

Is war more likely between populations that differ in terms of ethnicity, culture and historical legacies? Several commentators have argued that there is a general tendency towards violent confrontation between populations that are culturally and ethnically distant. For example, Maynes (1993, p. 5) wrote: "Animosity among ethnic groups is beginning to rival the spread of nuclear weapons as the most serious threat to peace that the world faces." Bremer (2000, p. 27), referring to evidence from social psychology, 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 between groups should be associated with "war and plunder," while societies that are ethnically and culturally related would tend to fight less with each other.

In this paper we present new empirical findings on the determinants of international conflict that strongly support the opposite view: populations that are more closely related are significantly more likely to engage in international conflict, even after controlling for a variety of measures of geographic proximity and other factors that affect conflict, including measures of trade and democracy. We explain this surprising result with an economic model of relatedness and conflict. The starting point is that populations inherit preferences with variation from their ancestors. On average, populations that share more recent common ancestors are also more similar in preferences. Similarity in preferences has very different implications for conflict depending on whether populations fight over control of goods that are *rival* in consumption and goods that are *non-rival* (public goods). Similarity is associated with a higher likelihood of conflict over rival goods, while more similar populations are less likely to fight over the control of public goods.

The central idea of this paper can be illustrated with a simple example. Consider two people in a room with two sandwiches: a chicken sandwich and a ham sandwich. People who share more similar preferences are more likely to want the same kind of sandwich, and possibly to fight over it, while people with more diverse preferences are more likely to be happy with different sandwiches. In contrast, suppose that there is a television set in the room, which both individuals must share. Each can watch television without reducing the other person's utility from watching, but they may disagree over which show to watch and fight over the remote control (public good). In this case, people with more similar preferences are less likely to fight, because they can agree on the same show. Since people who are more closely related tend to have more similar preferences on average - as they are passed along with variation from generation to generation - we expect that relatedness should be associated with more conflict when people fight over rival goods, and with less conflict when people fight over public goods.

What does this story have to do with international conflict? Historically, international conflict is mostly about control of rival and excludable goods, such as territories, cities and natural resources. The view that international conflict is mostly about territories and resources is emphasized, for instance, by Caselli, Morelli and Rohner (2012), who cite the results in Tir et al. (1998) and Tir (2003) that 27% of all territorial changes between 1816 and 1996 involve full-blown military conflict, and 47% of territorial transfers involve some level of violence.[1] In principle, sovereign states can fight over public goods (i.e., policies to deal with international terrorism or climate change) but historically such conflicts are a small subset of international disputes, and, more importantly, they are usually interlinked with control over rival resources. Conflicts over public goods are more likely to emerge among groups that belong to the same political jurisdictions and therefore share non-rival and non-excludable goods and policies by institutional design. Therefore, our framework has a straightforward implication for international conflict: relatedness should be associated with a higher probability of interstate war. In contrast, relatedness is likely to have ambiguous effects on intrastate conflict, because, within states, different groups tend to fight over a complex and variable mix of rival goods (i.e., natural resources) and non-rival goods (control over government type and policies).

To measure relatedness between populations, we use information on human genetic distance - a summary statistic of very long-run historical relatedness between populations. Genetic distance measures the difference in gene distributions between two populations, where the genes under considerations are neutral: they change randomly and independently of selection pressure. Most random genetic change takes place regularly over time, acting as a molecular clock (Kimura, 1968). Consequently, genetic distance measures 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

[1] They also cite Weede's (1973, p. 87) view that "the history of war and peace is largely identical with the history of territorial changes as results of war and causes of the next war."

[2] Specifically, we use measures of F_{ST} distance between human populations from Cavalli-Sforza et al. (1994).

siblings are more closely related than two cousins because they share more recent common ancestors - their parents rather than their grandparents. Since a very large number of characteristics - including cultural traits - are transmitted across generations over the long run, genetic distance provides a comprehensive measure of long-term distance in such traits across populations.

This paper's main empirical result is that genetic distance significantly reduces the risk of conflict, and the effect is substantial in magnitude. Populations that are more closely related are more likely to engage in interstate conflict and wars, even after controlling for several geographic variables, measures of linguistic and religious distance, and other factors that affect interstate conflict, including trade and democracy. A one standard deviation increase in genetic distance between two populations 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 possible endogeneity issues due to post-1500 migrations and population mixing. In a nutshell, from a long-term world-wide perspective, issues of war and peace are (unhappy) family matters.[3]

The negative effect of genetic distance holds when controlling for measures of geography (contiguity, geodesic distance, latitudinal and longitudinal differences, and other measures of geographic barriers). The paramount effect attributed by some scholars to geographic proximity (i.e., Gleditsch and Singer, 1975, Vasquez, 1995) may in part be due to its correlation with cultural and historical relatedness. Once genetic distance is taken into account, geographic variables have smaller effects, although they remain significant. In addition, the effect of genetic distance is robust when accounting for differences in military capabilities and income per capita across countries. We also control for other measures of cultural similarity, such as religious and linguistic distance. The effect of genetic distance is robust to such controls, and the effects of linguistic distance and religious

The measure F_{ST} was first suggested by the great geneticist and statistician Sewall Wright (1950). Interestingly, Sewall was the older brother of Quincy Wright, the professor of international law who pioneered empirical research on conflict (Wright, 1942). According to Singer (2000): "The story has it that [Sewall] admired Quincy's scholarship and his preoccupation with the scourge of war but lamented the lack of methodological rigor in his work and thus introduced him to the scientific method - hence the fifteen-year project that culminated in the monumental *Study of War* (1942)." We hope that the Wright brothers would appreciate our joining their two lines of research in a study titled "War and Relatedness." The Wrights were a truly remarkable family. As explained in Stock and Trebbi (2003), Sewall and Quincy's father Philip Wright was the inventor of instrumental variables regression.

[3] We apologize to Leo Tolstoy for the double plagiarism.

distance on international conflict are in line with the main predictions of the model: cultural similarity increases the probability of conflict. In particular, the fact that religious distance reduces the likelihood of war would be hard to rationalize within a clash-of-civilizations view (Huntington, 1993), but is consistent with the predictions of our framework. It is also worth noting that the effect of relatedness on conflict holds not only for the whole sample (1816-2001), but also by historical sub-periods, including for the post-Cold-War period (1990-2001).

Interesting results also emerge when adding measures of trade and democracy, to capture the central predictions of liberal peace theory: extensive bilateral trade links and the extent of democracy among countries in a pair should reduce their propensity to go to war.[4] Not only are the effect of relatedness robust to controlling for trade and democracy variables, but the effects of trade and democracy on conflict hold even after controlling for relatedness.[5] We are therefore able to address one of the most important criticisms of the empirical literature on this subject: observers who believe that culturally related countries fight less with each other have often questioned whether there is a direct causal link going from trade and democracy to lower conflict, on the ground that culturally more similar societies also tend to trade more with each other and to share more similar political arrangements (such as democratic regimes). Following this reasoning, the observed low level of conflict might not be the direct effect of trade and democracy, but rather the outcome of deeper cultural similarities (for discussions of this debate see, for example, Schneider, Barbieri and Gleditsch, 2003). In contrast, our estimates provide strong evidence that the premise that closely related populations fight less with each other is incorrect, and hence cannot account for the pacifying effects of bilateral trade and democracy. In sum, our findings validate the liberal view concerning the pacifying effects of trade and democracy, which continue to hold when controlling for relatedness.

This paper is the first, to our knowledge, to study the relationship between genetic distance

[4]The liberal peace view that trade and democracy should reduce the risk of war goes back to Montesquieu (1748) and Kant (1795). Contributions on the empirics of trade and conflict include Polacheck (1980), Oneal and Russett (1999a, 1999b), Barbieri (2002), and Martin, Mayer and Thoenig (2008). On the democratic peace hypothesis see, for instance, Maoz and Russett (1993), Bueno de Mesquita et al. (1999), Gowa (2000), and Levy and Razin (2004).

[5]Our empirical analysis of the effects of trade on conflict builds closely on Martin, Mayer and Thoenig (2008), and confirms their finding that bilateral trade reduces the risk of conflict between a pair of states, but multilateral trade increases the probability of conflict.

and the likelihood of interstate conflicts and wars.[6] It is part of a small but growing empirical literature on the connections between long-term relatedness and societal outcomes. In particular, while human genetic distance is not commonly used in the social sciences, recent work has pointed out to its usefulness and predictive power in economics and related areas. Spolaore and Wacziarg (2009) document the relation between genetic distance and differences in income per capita across countries, and provide an economic interpretation in terms of barriers to the diffusion of economic development from the world technological frontier. Desmet et al. (2011) find a close relationship between genetic distance and cultural differences measured by the World Values Survey, which supports our interpretation of genetic distance as a broad measure of differences in intergenerationally-transmitted traits, including cultural characteristics and preferences.[7] Our paper is thereby related to the evolutionary literature on cultural transmission of traits and preferences (Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985; Richerson and Boyd, 2004; for economic analyses of cultural transmission, see Bisin and Verdier, 2000, 2001) and to the growing empirical literature on the deep historical roots of economic and political outcomes (recent contributions include Ashraf and Galor, 2010, 2011; Putterman and Weil, 2010; and Michalopoulos and Papaioannou, 2010 and 2011, discussed among others in Spolaore and Wacziarg, 2012). Finally, this paper is part of a large and growing literature using formal theory and systematic empirical analyses to provide insights into the determinants of wars.[8]

[6]There are few formal or empirical analyses of the relations between war and genetic variables. Contributions by economists are Hirshleifer (1998), who provided a theoretical discussion of the evolutionary motives for warfare, including the "affiliative instinct" (partially related to the primordialist view), and, more recently, Bowles (2009), who studies whether warfare among ancestral hunters-gathers may have affected the evolution of group-beneficial behavior.

[7]Desmet et al. (2011) find that European populations that are genetically closer give more similar answers to a broad set of 430 questions about norms, values and cultural characteristics, included in the 2005 World Values Suvey sections on perceptions of life, family, religion and morals. They also find that the correlation between genetic distance and differences in cultural values remains positive and significant after controlling for linguistic and geographic distances.

[8]Economic analyses of conflict and wars include Schelling (1960), Boulding (1962), and, among the more recent contributions, Garfinkel (1990), Skaperdas (1992), Hess and Orphanides (1995, 2001), Alesina and Spolaore (2005, 2006), Martin, Mayer and Thoenig (2008), Jackson and Morelli (2007), Yared (2010), Caselli, Morelli and Rohner (2012) among many others. Garfinkel and Skaperdas (2007) provide an overview of the economics literature on conflict. Of course there exists a much larger literature on this topic in history (e.g., Blainey, 1988 and Ferguson, 2006) and in political science (e.g., Bueno de Mesquita and Lalman, 1992, Fearon, 1995, Powell, 1999 and Russett and

While the theoretical and empirical focus of our paper is on interstate conflict, a natural extension would be to study the effects of relatedness on intrastate conflict. As far as we know, there is no study that directly explores the effects of genetic relatedness on civil conflict, perhaps because of data limitations in measuring genetic relatedness between groups with the level of detail that a within-country study would require. Of course, there is a huge literature that considers the effects of various measures of ethnic divisions on intrastate conflict, with sometimes conflicting results. For instance, Fearon and Laitin (2003, p. 82) in their influential study of the effects of ethnic fractionalization on civil conflict, concluded that the observed "pattern is thus inconsistent with [...] the common expectation that ethnic diversity is a major and direct cause of civil violence." In contrast, the literature on ethnic polarization, pioneered by the theoretical contribution of Esteban and Ray (1994) and the empirical work of Montalvo and Reynal-Querol (2005), leads to different conclusions. In a recent empirical study of ethnicity and intrastate conflict, Esteban, Mayoral and Ray (2012), building on the previously cited papers, find that when civil conflict is mostly over public goods, ethnolinguistic polarization leads to more conflict. This finding is consistent with our hypothesis that less closely related groups are more likely to fight over the control of public goods. More broadly, since civil conflicts are about a complex mix of disputes over rival and public goods, our findings are consistent with the lack of consensus found in the literature concerning the role of ethnic divisions for civil conflict.

The rest of the paper is organized as follows. Section 2 presents a model of relatedness and conflict (extensions are included in the Appendix). Section 3 introduces the data and methodology. Section 4 discusses the empirical findings. Section 5 concludes.

2 A Model of Relatedness and Conflict

In this section we present a theoretical framework linking genetic distance, intergenerationally-transmitted preferences, and the probability of conflict between societies. First, we model the transmission of preferences over time with variation across populations. Populations that are more closely related (i.e., at a smaller genetic distance) tend to have more similar preferences. Second, we model conflict over rival goods. Conflict is more likely to arise when different populations strongly

Oneal, 2001). Systematic empirical work on interstate conflict was pioneerd by Wright (1942), Richardson (1960) and Singer (1972). For discussions of the empirical literature on the correlates of war see Vasquez (2000) and Schneider, Barbieri and Gleditsch (2003).

care about the same rival goods and resources. The analysis leads to a testable prediction: the probability of conflict over rival goods between two societies should be inversely related to their genetic distance. In Section 4, we show that the empirical evidence on international conflict strongly supports this prediction. Finally, we discuss how the effect of relatedness on conflict would change if the dispute were about control of goods that are non-rival in consumption (public goods).

2.1 Relatedness and the Transmission of Preferences

We first present a simple model of the intergenerational transmission of preferences over the very long run. Consider three periods (o for origin, p for prehistory, and h for history). In period o there exists only one population (population 0). In period p the original population splits into two populations (1 and 2). In period h each of the two populations splits into two separate populations again (population 1 into 1.1 and 1.2, and population 2 into 2.1 and 2.2), as displayed in Figure 1. In this setting, the genetic distance $d_g(i, j)$ between population i and population j can be simply measured by the number of periods since they were one population:

$$d_g(1.1, 1.2) = d_g(2.1, 2.2) = 1 \tag{1}$$

and

$$d_g(1.1\ ,\ 2.1) = d_g(1.1\ ,\ 2.2) = d_g(1.2\ ,\ 2.1) = d_g(1.2\ ,\ 2.2) = 2 \tag{2}$$

These numbers have an intuitive interpretation: populations 1.1 and 1.2 are sibling populations, sharing a common parent ancestor (population 1), while populations 2.1 and 2.2 are also sibling populations, sharing a different common parent ancestor (population 2). In contrast, populations 1.1 and 2.1, for example, are cousin populations sharing a common grand-parent ancestor (population 0).

For simplicity, preferences are summarized by two types (A and B). At time o, the ancestral population 0 is either of type A or of type B. For analytical convenience and without loss of generality, we assume that population 0 is of type A with probability 1/2 and of type B with probability 1/2.[9] Populations inherit preferences from their ancestors with variation - a population i' descending from a population i will have preferences of the same type as their parent population i with probability μ, and of the other type with probability $1 - \mu$.

[9] The qualitative results would not change if we were to assume that the ancestral population is of type A with probability 100% or of type B with probability 100%.

We capture the fact that populations inherit preferences from their ancestors by assuming $\mu > 1/2$ and the fact that there is variation (inheritance is not perfect) by assuming $\mu < 1$.[10] Then, on average, populations at a smaller genetic distance from each other will tend to be more similar in preferences. For instance, the probability that two sibling populations (e.g., 1.1 and 1.2) have identical types is

$$F(\mu) = \mu^2 + (1-\mu)^2 \tag{3}$$

while the probability that two cousin populations (e.g., 1.1 and 2.1) have identical types is

$$G(\mu) = \mu^4 + 6\mu^2(1-\mu)^2 + (1-\mu)^4 \tag{4}$$

It can be easily shown that[11]

$$F(\mu) > G(\mu) \qquad \text{for } \frac{1}{2} < \mu < 1 \tag{5}$$

which implies:

Proposition 1

The probability that two populations are of the same type is decreasing in genetic distance.

This result plays a key role in our analysis of conflict below.

[10] At $\mu = 1/2$, each population would have equal chances of being of either type, independently of the parent population's type, while at $\mu = 1$, each population would be of the same type as their ancestors with 100% probability.

[11] By dividing both $F(\mu)$ and $G(\mu)$ by μ and rearranging terms, the inequality $F(\mu) - G(\mu) > 0$ can be re-written equivalently as

$$2 - 10\mu + 16\mu^2 - 8\mu^3 \equiv f(\mu) > 0$$

It is immediate to verify that the above inequality holds, given that $f(\frac{1}{2}) = f(1) = 0$ and the derivative

$$f'(\mu) = 2(-5 + 16\mu - 12\mu^2)$$

is strictly positive for $1/2 < \mu < 5/6$, zero at $\mu = 5/6$, and negative for $5/6 < \mu \leq 1$.

2.2 Conflict

Consider two populations (i and j), each forming a sovereign state.[12] Suppose that sovereign state i is in control of a valuable prize of type t, from which it obtains the following benefits b_i:

$$b_i = (1 - |t - t_i^*|)R \tag{6}$$

where t_i^* denotes state i's ideal type, and $R > 0$ is the size of the prize. If the prize is of type A, $t = t_A$, and if it is of type B, $t = t_B$. Without loss of generality, we assume that the prize is of type A with probability 1/2 and of type B with probability 1/2. State i's ideal type is also equal to either t_A or t_B. We assume that the state benefits from controlling the prize even if it is not of its favored type, that is

$$|t_A - t_B| < 1 \tag{7}$$

The prize can be interpreted as any valuable good which can be controlled by a sovereign state - e.g. natural resources, land, cities, trade routes, colonies, protectorates, etc. (we return to the interpretation of the model below, when we discuss possible extensions). Sovereign state j also values the prize, and would gain positive benefits if it could control the prize. State j's benefits $b_j$from controlling the prize are

$$b_j = (1 - |t - t_j^*|)R \tag{8}$$

State j can try to obtain control over the prize by challenging state i - that is, state j can take two actions: "challenge" state i (C) or "not challenge" (NC). If state j chooses action NC, state i keeps full control over the prize, and obtains a net utility equal to b_i, while state j obtains net benefits equal to zero. If state j challenges state i for the possession of the prize, state i can respond either with "fight" (F) or "not fight" (NF). If state i does not fight, state j obtains control of the prize, and net benefits equal to b_j, while state i obtains net benefits equal to zero.

If state i decides to fight in response to the challenge, a war takes place.[13] When a war occurs

[12] For simplicity, we assume that each state is a unified agent, formed by one population with homogeneous preferences. In our model, we abstract from the possibility that states may include mixed populations with different preferences. However, for the empirical analysis we take into account population heterogeneity within states when computing genetic distance. In this theoretical analysis we also abstract from the possibility that the state may be controlled by a non-democratic Leviathan that pursues objectives different from those of the whole population. In the empirical section we control for measures of democracy.

[13] In the Appendix (A.1) we present an extension in which peaceful bargaining is possible, in alternative to war, when state j challenges and state i responds to the challenge.

(i.e., when actions $\{C, F\}$ are taken), the probability that state i wins, denoted by π_i, is a function of the two states' relative military capabilities (denoted respectively by M_i and M_j):[14]

$$\pi_i = \frac{M_i}{M_i + M_j} \tag{9}$$

while the probability that state j wins the war is obviously $1 - \pi_i$. Ex ante, each state obtains an expected utility respectively given by

$$U_i = \pi_i b_i - c_i \tag{10}$$

$$U_j = (1 - \pi_i) b_j - c_j \tag{11}$$

where $c_i > 0$ and $c_j > 0$ denote the respective costs of going to war. The extensive form of the game is illustrated in Figure 2.

It is immediate to show that:

Lemma

War is a sub-game perfect equilibrium if and only if $\min\{U_i, U_j\} \geq 0$. *War is the unique sub-game perfect equilibrium when* $\min\{U_i, U_j\} > 0$.[15]

2.3 War and Relatedness

We now investigate how similarity in preferences between the two states affect the probability of war. To simplify the analysis, we assume equal capabilities ($M_i = M_j = M$) and costs ($c_i = c_j = c$). Let $P(i, j)$ denote the probability of a war between state i and state j.

[14] This is an instance of "ratio" contest success function. In general, the literature on the technology of conflict assumes that the probability of success is a function of either the ratio or the difference between military capabilities (for a general discussion, see Garfinkel and Skaperdas, 2007). The choice of specification in this paper is inconsequential because we treat military capabilities as exogenous. A straighforward extension would be to endogenize military capabilities. The extension could strengthen the link between relatedness and probability of conflict, insofar as states with similar preferences might face more similar incentives to invest in military capabilities, all other things being equal. We do not pursue the extension here, but we control for differences in military capabilities in the empirical section.

[15] When $U_i > 0$ and $U_j = 0$, two sub-game perfect equilibria exist: $\{C, F\}$ and $\{NC, F\}$. When $U_i = 0$ and $U_j > 0$, there are also two sub-game perfect equilibria: $\{C, F\}$ and $\{C, NF\}$. When $U_i = U_j = 0$, three equilibria may occur: $\{C, F\}$, $\{C, NF\}$ and $\{NC, F\}$. When $\min\{U_i, U_j\} < 0$ the only sub-game perfect equilibria are peaceful. If $U_i < 0$, the only sub–game perfect equilibrium is $\{C, NF\}$. If $U_i > 0$ and $U_j < 0$, the only sub-game perfect equilibrium is $\{NC, F\}$. Finally, when $U_i = 0$, and $U_j < 0$ there are two (peaceful) equilibria: $\{NC, F\}$ and $\{C, NF\}$.

Clearly, a war would *never* occur ($P(i,j)=0$) if each state's expected utility from going to war is negative even when the prize is of its preferred type. This would happen at a very high cost of war:

$$c > \frac{1}{2}R \tag{12}$$

In contrast, a war would *always* occur ($P(i,j)=1$) if each state's expected utility from going to war is positive even when the resource is not of its favored type. This would happen at a very low cost of war:

$$c < \frac{1}{2}R(1-|t_A - t_B|) \tag{13}$$

Therefore, we will focus on the more interesting case when war may occur with probability between 0 and 1 ($0 < P(i,j) < 1$), which happens when the cost of war takes on an intermediate value[16]:

$$\frac{1}{2}R(1-|t_A - t_B|) < c < \frac{1}{2}R \tag{14}$$

Under these assumptions, a war will occur if and only if the two states have the same preferred type, and that type is equal to the type of the prize under dispute - that is, $t_i^* = t_j^* = t$. If the two states had always identical preferences, the probability of a war would be 1/2.[17] In contrast, if the preferences of each state were independently distributed, with each state having a 50% chance of preferring type A to type B (and vice versa), the probability of war would be 1/4.[18]

In general, for $1/2 < \mu < 1$, the expected probability of war between states i and j would depend on the degree of relatedness (genetic distance) of their populations. For two states i and j with $d_g(i,j) = 1$ - i.e., states formed by sibling populations - the probability that both states' type is the same as the prize under dispute is half the probability that both states have the same preferences, that is

$$P\{i,j \mid d_g(i,j)=1\} = \frac{F(\mu)}{2} = \frac{\mu^2 + (1-\mu)^2}{2} \tag{15}$$

By the same token, for states such that $d_g(i,j) = 2$ - i.e., states formed by cousin populations -, the probability that both states' type is equal to the type of the prize is

$$P\{i,j \mid d_g(i,j)=2\} = \frac{G(\mu)}{2} = \frac{\mu^4 + 6\mu^2(1-\mu)^2 + (1-\mu)^4}{2} \tag{16}$$

[16] To simplify the analysis, we do not consider the knife-edge cases $c = \frac{1}{2}R$ and $c = \frac{1}{2}R(1-|t_A - t_B|)$, when it's possible that $\min\{U_i, U_j\} = 0$, implying that one or both states may be indifferent between war and peace, and multiple equilibria may trivially occur, as detailed in the footnote to the Lemma.

[17] This would occur, for instance, if preferences were transmitted without variation across generations: $\mu = 1$.

[18] This would occur, for instance, if preferences were transmitted purely randomly across generations: $\mu = 1/2$.

As already shown in the previous section, $F(\mu) > G(\mu)$ for all $1/2 < \mu < 1$. Therefore, it immediately follows that

$$P\{i, j | d_g(i, j) = 1\} > P\{i, j \mid d_g(i, j) = 2\} \tag{17}$$

which we can summarize as our main result:

Proposition 2

States with more closely related populations (smaller genetic distance) are more likely to go to war with each other.

An illustration of the model can be provided with a simple spatial example. Assume that space is unidimensional. Three states divide the territory among themselves as in Figure 3, with the border between state i and state j at point x, and the border between state j and state i' at point y. Assume that state i and state j are of type A, and state i' is of type B. The parameters are such that equation (14) is satisfied. Now, consider the territory between x' and x. If that territory is of type B, state j will not challenge state i for its possession, but if that territory is of type A, a war will occur. In contrast, consider the territory between y and y'. If that territory is of type B, state j will not challenge state i' for its possession, while if it is of type A, state j will challenge state i', and state i' will surrender it peacefully. In either case, no conflict will occur.

This example illustrates how the probability of conflict between states in similar geographical settings varies because of preferences over the prize: states with more similar preferences are more likely to go to war with each other, other things being equal. In this example, the prize is a contiguous territory, but similar effects would hold for control over non-contiguous territories (colonies, protectorates, ports and harbors along trade routes), or over other rival issues about which states may care with different intensities (for instance, monopoly rights over trade, fishing or other valuable sources of income in specific waters or regions). History is abundant with examples of populations that fought over specific rival goods (territories, cities, religious sites) because they shared a common history and common preferences, inherited with variation from their ancestors. For instance, genetically close populations (Jews and Arabs) who share similar preferences over Jerusalem have fought and continue to fight over the control of that rival good. In general, we can expect that populations may share more similar preferences over specific types of land and resources because they have inherited similar tastes and demand functions (as in the example about Jerusalem), or because they have inherited similar technologies and methods of production, or both.[19] Preferences for specific rival goods can be inherited culturally or biologically, or as a result of dual inheritance (gene-culture coevolution) - for example, those populations who inherited the gene mutation allowing the digestion of milk by adults, along with the ability to domesticate and exploit milk-producing animals, historically might have cared much more about cows and pastures - and fought over them - than populations who did not share those inherited traits.[20]

2.4 Conflict over Public Goods

In our basic model the prize is a rival and excludable good: either one or the other state obtains full control, and the population in the state without control receives no net benefit. How would our results change if the prize were a public good (non-rival and non-excludable in consumption)? Then, state j would obtain some external benefits when state i is in control of the good, and vice versa. In itself, this extension would only reduce the likelihood of war, because the externalities would reduce the gap in utility between controlling and not controlling the good.

However, the implications would change dramatically if we also allowed the state in control to select the characteristics or type of the public good. As we show formally in Appendix A.2 this modification reverses the main result of our basic model.[21] The intuition is straightforward: Suppose conflict is not about control of the public good *per se*, but about determination of its type. Then, more closely related populations, sharing more similar preferences about the characteristics of the public good, are less likely to engage in conflict. In contrast, populations that are historically and culturally more distant tend to disagree more over the type of public good. Therefore, similarity in preferences over public goods could partly or totally offset the effects of similarity in preferences over rival goods, depending on whether the conflict is mostly about rival goods or public goods.

Conflicts over public goods are more likely to emerge among groups that belong to the same

[19] In principle, similarities in technology could affect the probability of conflict not only by affecting preferences over rival goods, but also, more directly, by affecting military capabilities (more similar populations may be more similar in military technologies and hence capabilities, other things being equal). As already mentioned, in the empirical section we control for differences in military capabilities, and continue to find an effect of genetic distance on conflict. In any case, we find that differences in measured capabilities are not correlated with genetic distance in the data.

[20] For a discussion of inheritance mechanisms (biological, cultural and dual) and their relevance for economic and politial outcomes see Spolaore and Wacziarg (2012).

[21] A generalization of the framework that encompasses both models (conflict over pure rival goods and conflict over pure public goods) as special cases is presented in Appendix A.3.

political jurisdictions and therefore share non-rival and non-excludable goods and policies by institutional design.[22] Therefore, conflict over public goods and policies is likely to play an important role in many (but not all) civil conflicts. This observation is consistent with theoretical and empirical work associating ethnic polarization, a measure that captures distance between groups within a country, with conflict over public goods. Of particular note is a recent empirical study of ethnicity and intrastate conflict by Esteban, Mayoral and Ray (2012), building on theoretical work by Esteban and Ray (2011). They use measures of ethnolinguistic polarization using linguistic distance between groups, and find empirically that linguistic polarization increases intrastate conflict over public goods.[23] Such effects of distance and polarization, albeit linguistic rather than genetic, are entirely consistent with our hypothesis that less closely related groups are more likely to fight over the control of public goods.[24]

In contrast to the case of civil conflict, the importance of this public-goods effect is likely to be much lower for interstate conflict. Even though disagreements about the provision of public goods and policies may also emerge among sovereign states (i.e., how to fight international terrorist threats, global climate change, or financial instability), historically interstate militarized conflicts have been mostly about control of rival and excludable goods, such as territories, cities, and natural resources. In our empirical analysis we focus on the determinants of interstate conflict, where the basic model is likely to capture the main forces at work, leaving the study of the more complex links between relatedness and civil conflict for further research.

[22]See Alesina and Spolaore (1997) for a discussion of the relation between diversity of preferences over public goods and the costs of forming more heterogeneous states, and Desmet et al. (2011) for an analysis linking such heterogeneity costs to historical relatedness between populations. See also Spolaore (2008, 2012) for formal analyses of conflict in that context.

[23]Esteban, Mayoral and Ray (2012) use measures of linguistic distance, based on those in Fearon (2003) and Desmet, Ortuño-Ortíz and Wacziarg (2012).

[24]The theory in Esteban and Ray (2011) and Esteban, Mayoral and Ray (2012) also draws a distinction between public goods and private goods. In their framework, a central role is played by three indices, measuring polarization, fractionalization and cohesion. The weight of these indices in explaining conflict intensity depends on the particular nature of each conflict: when group cohesion is high, ethnic polarization increases conflict if the prize is public and fractionalization increases conflict if the prize is private. Our analysis and results are quite different from theirs. Our focus is on the effects of relatedness between populations, not on polarization and fractionalization within societies. More importantly, while our model predicts a positive effect of genetic distance on conflict over public goods, consistent with their results on polarization, our model also predict a *negative* effect of distance on conflict over private goods.

3 Data and Methodology

Our model predicts that the degree of relatedness between populations has a positive effect on their conflict propensities. In the remainder of this paper we examine empirically the determinants of bilateral conflict across states, focusing on the degree of relatedness between the populations of each pair of countries. We control for other determinants of bilateral conflict, in particular numerous measures of geographic distance.

3.1 Measuring Conflict

We use panel data on interstate conflict between 1816 and 2001 from the Correlates of War Project (www.correlatesofwar.org).[25] 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 of militarized conflict is equal to or greater than 3, zero otherwise. Since our main independent variable is time-invariant, our focus is mainly on 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 of interest is this binary indicator of conflict, denoted C_{ij}. We separately examine the determinants of the maximal intensity of conflict, as well as the determinants of full-blown war (corresponding to a pair having ever experienced a conflict intensity equal to 5). We also separately conduct an analysis of the determinants of C_{ijt}, i.e. exploiting the time dimension, in order to control for time-varying factors affecting conflict propensities.

3.2 Measuring Relatedness

To capture genealogical relatedness, 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. Genetic distance is a summary measure of differences in allele frequencies across a range of neutral genes (or *chromosomal loci*). The measure we use, F_{ST} genetic distance, captures the length of time since two populations became separated from each other. When two

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

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 genealogical relatedness, resulting from a molecular clock. The specific source for our data is Cavalli-Sforza et al. (1994), pp. 75-76.[26]

Our focus is on a set of 42 world populations for which there is data on bilateral genetic distance, computed from 120 neutral alleles. Among the set of 42 world populations, the maximum genetic distance is between Mbuti Pygmies and Papua New-Guineans ($F_{ST} = 0.4573$), and the minimum is between the Danish and the English ($F_{ST} = 0.0021$). The mean genetic distance among the 861 available pairs is 0.1338.

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 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 provides 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 $m = 1...M$ and country j is composed of populations $n = 1...N$.[27] Denote by s_{1m} 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:

$$FST_{ij}^{W} = \sum_{m=1}^{M} \sum_{n=1}^{N} (s_{im} \times s_{jn} \times d_{mn}) \tag{18}$$

where s_{km} is the share of group m in country k, d_{mn} is the F_{ST} genetic distance between groups

[26] Cavalli-Sforza et al. (1994) also provide data on Nei genetic distance, a measure that is different but highly correlated with F_{ST} distance. Our results are robust to using Nei distance rather than F_{ST} distance. Corresponding estimates are available upon request.

[27] That 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.

m and n. 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%), so for practical purposes it does not make a big difference which one we use. We will 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.[28]

The match of populations to countries pertains to the contemporary period, after the great migrations that followed the conquest of the New World. Hence, for instance, for the current period the plurality population in Australia is the English population. To address bias resulting from errors in matching populations to countries for the current period, as well as concerns that current genetic distance may be endogenous with respect to past wars, we also matched countries to their 1500 AD populations. For instance, 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.[29]

3.3 Summary Statistics

Table 1 and 2 provide basic statistics that give a general sense of the data and provide clues concerning the relationship between conflict and relatedness. The statistics pertain to a baseline sample of $13,575$ country pairs, based on 176 underlying countries. The underlying data is an unbalanced panel of $517,251$ observations with yearly observations from 1816 to 2001, but for the purposes of summary statistics the panel has been collapsed into a cross-section.[30] 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

[28] Our results are robust to using genetic distance between plurality groups rather than weighted genetic distance. The corresponding estimates are available upon request.

[29] Since we do not have detailed data on ethnic composition going back to 1500, the corresponding match only refers to plurality groups. Matching countries to populations for 1500 is more straightforward than for the current period, since Cavalli-Sforza et al. (1994) attempted to sample populations as they were in 1500. This likely reduces the extent of measurement error. The correlation between weighted genetic distance matched using current period populations and genetic distance between plurality groups as of 1500 is 0.723 in our baseline sample.

[30] Summary statistics computed from the uncollapsed panel (i.e. using C_{ijt} instead of C_{ij} as the measure of conflict) give a message similar to the cross-sectional ones, and are available upon request.

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 have ever experienced full-blown wars between 1816 and 2001, out of $13,175$ pairs. 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. While South Africa is characterized in our data as genetically distant from European populations due to the large African majority there, 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 computed in a way that works *against* finding a positive link between relatedness and conflict. 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 it is still the case that 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 wars *per se*: 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 very much when conditioning on geographic distance being large between the countries in 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, but to examine this hypothesis more formally we turn to regression analysis.

3.4 Empirical Specification

A more formal 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.

We consider two baseline methodologies. The main methodology we focus on is cross-sectional: we collapse the panel into a single cross-section where the dependent variable is the already defined indicator of whether a pair was ever in an interstate conflict between 1816 and 2001. Since our main independent variable of interest, F_{ST} genetic distance, is time invariant at the horizon of this study, it is a natural starting point to consider the determinants of whether a country ever had a conflict or a war over the 1816 to 2001 time period. The baseline cross-sectional regression specification is:

$$C_{ij} = \beta_1 X_{ij} + \beta_2 FST_{ij}^{W} + \eta_{ij} \tag{19}$$

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.

The second methodology is to make full use of the panel dimension. This allows us to control for time varying determinants of conflict, some of which (democracy, trade, income differences, among others) are important control variables that have given rise to important strands of the literature on the determinants of international conflict. The baseline panel specification is:

$$C_{ijt} = \gamma_1 X_{ijt} + \gamma_2 FST_{ij}^{W} + \varepsilon_{ijt} \tag{20}$$

where X_{ijt} contains all of the aforementioned geographic and colonial controls plus time varying measures such as a dummy variable representing whether both countries in the pair are democracies, whether they belong to an active military alliance, how many years they have been at peace with each other, and the number of other wars occurring in year t. The choice of controls in X_{ijt} closely follows the existing literature, particularly the contribution of Martin, Mayer and Thoenig (2008). A major difference is that we greatly augment the list of geographic controls compared to existing contributions, in an effort to identify separately the effects of geographic proximity from those of genealogical relatedness. It is important for our purposes to adequately control for geographic isolation as genetic distance and geographic isolation tend to be correlated (for instance the correlation between F_{ST} genetic distance and log geodesic distance in our baseline sample is 0.434). It is important to note, however, that the correlation between genetic distance and geographic distance is far from perfect. In particular, the massive populations movements that followed the discovery of the New World (both due to the European conquests and to the slave trade) have served to greatly reduce the correlation between genetic distance and geographic distance.

Equations (19) and (20) are estimated using probit. For the panel specification, we cluster standard errors at the country-pair level. 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.

4 Empirical Results

4.1 Baseline Cross-Sectional Estimates

Table 3 presents baseline estimates of the coefficients in equation (19). We start with a univariate regression (column 1), showing a very strong negative relationship between genetic distance and the incidence of militarized conflict. The magnitude of this effect is large, as a one standard deviation change in genetic distance (0.068) is associated with a 68.81% decline in the percentage probability of to 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.

Column (2) introduces eight measures of geographic distance, plus two measures of colonial past.[31] 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,

[31] We also included various measures of climatic similarity within country pairs, using Koppen-Geiger codings of climate. The idea was that similar countries might seek to conquer countries with similar geographies. The inclusion of these variables did not lead to discernible changes in the effect of genetic distance (results are available upon request).

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 via naval projection of force (we find, in fact, that pairs composed of 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.[32] However, this effect remains negative and highly 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. There are two main issues. The first issue is measurement error stemming from imperfect matches of genetic groups to current populations and countries, leading to probable attenuation bias. The second issue is reverse causality. 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. However, population 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 (Cavalli-Sforza, Menozzi, and Piazza, 1994).[33]

To address endogeneity and measurement error, we instrument for modern genetic distance using genetic distance between populations as they were in 1500. Genetic distance in 1500 is unlikely to be causally affected by conflicts between 1816 and 2001. Moreover, matching countries

[32] Proceeding sequentially, we found that adding these controls one by one progressively reduced the effect of genetic 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.

[33] For instance, Opennheimer (2006) argues that the genetic composition of the English population is dominated by that of the populations, from the Iberic Peninsula and Central Europe, that populated the British Isles after the end of the Ice Age. Subsequent major invasions, from the Romans, the Anglo-Saxons, the Normans, etc., contributed very little to the English gene pool: the genetic composition of the English was basically settled 3,000 years ago.

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 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.

The results using 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. The higher effect of genetic distance under IV is likely to reflect lower prevalence of measurement error, since arguments about reverse causality would suggest that instrumenting should reduce the effect of genetic distance. To adopt a conservative approach, in the rest of the analysis we will 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). We redefine the dependent variable as a binary indicator of war, i.e. a dummy variable equal to one if the pair ever experienced a conflict of intensity equal to 5 (corresponding to conflicts with more than 1,000 total battle deaths), over the sample period. Only 2.09% of the country pairs in our sample ever experienced a full-blown war, so-defined, between 1816 and 2001. 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.[34]

[34] Additionally, 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 Henehan, 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-terrorial conflicts. These empirical results are available upon request.

4.2 Estimates Across Time and Space

To examine if the negative effect of relatedness on conflict across various geographic locations, we break down the sample across space, by continent. The goal is in part to establish whether the overall result might be driven by a specific continent. Additionally, the fact that many conflicts occur within continents might be driving the negative effects of relatedness in the worldwide sample, since intracontinental genetic distance is typically much smaller than cross-continental genetic distance. Finally, geographic barriers to conflict across continents are much larger than within continents, so looking for within-continent effects of genetic distance on conflict propensities is a way to further address the possibly confounding effects of geographic factors. To do so, we isolate pairs of countries that belonged to the same continent - defined as Europe, Asia, Africa and the Americas, and examine the determinants of conflicts among those pairs separately.[35] The results appear in Table 4.

We find 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 (column 2), we have the advantage of observing a separate, more detailed matrix of F_{ST} genetic distance.[36] The results are particularly striking in this subsample: despite the paucity of observations (only 291 country pairs), the effect of genetic distance remains negative and significant at the 1% level, and its standardized magnitude (38.4%) is over 60% larger than in the worldwide sample (23.8%). The result is significant because 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, and is thus much smaller in magnitude than genetic distance in the worldwide sample. To be able to identify a large positive effect of relatedness on conflict propensities among populations that are closely related

[35] In our baseline sample of 13,175 pairs, the number of pairs having experienced intracontinental interstate conflicts between 1816 and 2001 is 112 in Asia (out of 866 pairs), 75 in Africa (out of 1,048 pairs), 68 in the Americas (out of 581 pairs) and 71 in Europe (out of 291 pairs). There were no conflicts among the 27 country pairs located in Oceania.

[36] Estimates 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, showing the robustness of the effect of genetic distance to the inclusion of additional microgeography controls and sample splits by time periods, are available upon request.

historically reinforces the robustness of our main result.

The strong effects found within Europe raise the possibility that the worldwide results are driven by Europe. To address this possibility, we isolate every pair of countries that did not involve at least one European country. Results are presented in column (3) of Table 4. We find that, while slightly smaller in magnitude than for the full sample, the negative effect of genetic distance remains large and statistically significant. Thus, our worldwide results were not driven only by conflicts involving European countries. Overall, the regional breakdown suggests that the negative effect of relatedness on war is remarkably consistent across space.

We next examine whether relatedness affects conflict differently across time. To do so, we define dummy variables for whether a country pair was ever at war during a specific subperiod.[37] Results are presented in Table 5. We find again 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 20^{th} century, in particular the two World Wars. Focusing on the 20^{th} 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. In fact, our finding holds strongly even after the end of the Cold War (column 6), despite the relatively small number of pairs involved in conflicts during this period (only 218). The effect of genetic distance is negative and statistically significant whatever the subperiod under consideration, and the magnitude is large particularly for the subperiods spanning the 20^{th} Century.

4.3 Adding Linguistic and Religious Distance

While genetic distance is a precise and continuous measure of the degree of relatedness between populations and countries, other measures exist. The existing literature on interstate conflict has examined linguistic and religious ties in an effort to tell apart primordialist theories of conflict from instrumentalist theories (Richardson, 1960, Henderson, 1997). Thus, it is important to evaluate whether these variables trump genetic distance, and more generally how their inclusion affects our main coefficient of interest. Linguistic relatedness is associated with genetic relatedness because,

[37] These 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).

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).[38] Religious beliefs, also transmitted intergenerationally, are one type of human traits that can affect conflict. 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.

Prior to discussing the results, we describe how these measures were constructed. To capture linguistic distance, we use the data and approach in Fearon (2003), making use of linguistic trees from Ethnologue to compute the number of common linguistic nodes between languages in the world, a measure of their linguistic similarity (the linguistic tree in this dataset involves up to 15 nested classifications, so two countries with populations speaking the same language will share 15 common nodes).[39] Using data on the distribution of each linguistic group within and across countries, from the same source, we compute a measure of the number of common nodes shared by languages spoken by plurality groups within each country in a pair. We also compute a weighted measure of linguistic similarity, representing the expected number of common linguistic nodes between two randomly chosen individuals, one from each country in a pair (the formula is analogous to that of equation 18).[40] Following Fearon (2003), we transform these measures so that they reflect

[38] 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 very distinct populations genetically. See Spolaore and Wacziarg (2009) for an extensive discussion of these points.

[39] As an alternative, we used a separate measure of linguistic distance, based on lexicostatistics, from Dyen, Kruskal and Black (1992). This is a more continuous measure than the one based on common nodes, but it is only available for countries speaking Indo-European languages. It captures the number of common meanings, out of a list of 200, that are conveyed using "cognate" or related words. Summing over the 200 meanings, a measure of linguistic distance is the percentage of non-cognate words. Using the expected (weighted) measure of cognate distance led to effects of genetic distance very similar to those obtained when controlling for the Fearon measure, albeit on a much smaller sample of countries speaking Indo-European languages. These results are available upon request.

[40] The two measures deviate from each other whenever a country includes populations speaking different languages. Using the measure based on the plurality language or the weigthed measure did not make any difference for our results. As we did for genetic distance, we focus on weighted measures.

linguistic distance (LD) rather than similarity, and are bounded by 0 and 1:

$$LD = \sqrt{\frac{(15 - \#\ \text{Common Nodes})}{15}} \tag{21}$$

To measure religious distance we follow an approach based on religious trees, similar to that used for linguistic distance, using a nomenclature of world religions obtained from Mecham, Fearon and Laitin (2006). This nomenclature provides a family tree of World religions, first distinguishing between monotheistic religions of Middle-Eastern origin, Asian religions and "others", and further subdividing these categories into finer groups (such as Christians, Muslims and Jews, etc.). The number of common classifications (up to 5 in this dataset) is a measure of religious similarity. We match religions to countries using Mecham, Fearon and Laitin's (2006) data on the prevalence of religions by country and transform the data in a manner similar to that in equation (21), again computing plurality and weighted distances separately.[41]

Table 6 presents estimates of the effect of genetic distance on the propensity for interstate conflict when linguistic and religious distance 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.[42] 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.[43]

[41] Pairwise correlations between measures of genetic, linguistic and religious distances are generally positive, as expected, but not very large. For instance, the correlation between F_{ST} genetic distance and weighted linguistic distance is 0.201. Religious distance (weighted) bears a correlation of 0.449 with linguistic distance, and 0.172 with genetic distance.

[42] This 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).

[43] The estimated effects of religious and linguistic distance do not change much when genetic distance is excluded

4.4 Nonlinearities and Determinants of Conflict Intensity

In this subsection, we consider several extensions of our baseline specification. Our goal is to characterize whether relatedness may operate differently for different pairs of countries, and to investigate its effect on the intensity of conflict. To do so, we first look for interactive and nonlinear effects of genetic distance (Table 7). We then seek to evaluate the effect of genetic distance on the intensity of conflict, rather than on the binary indicator of whether a pair ever experienced a conflict (Table 8).

We first isolate countries that are non contiguous. In the baseline sample, 20.6% of country pairs having experienced conflicts (i.e. 153 out of 744 pairs) involve contiguous countries, and isolating pairs composed of non-contiguous countries is a further way to control for the possibility that geographic factors drive the effect of genetic distance (column 2 of Table 7). The standardized effect of genetic distance actually *rises* modestly, as a one standard deviation increase in genetic distance is associated with a 27.34% decrease in the mean probability of having experienced a conflict (versus 23.83% in the baseline regression). This reinforces our confidence that the effect is not driven by geographic factors or other possibly omitted factors specific to contiguous countries.

In columns (3) through (5) we add several interaction terms to the baseline specification. The effect of genetic distance does not appear quantitatively more or less pronounced for pairs that include a major power or for pairs that are geographically proximate (i.e. countries are either contiguous or separated by a distance less than 2,500 km). We find some evidence that the negative effect of genetic distance is reversed for pairs that are contiguous, although the proportion of contiguous pairs in the sample is so low (and the number of conflicts among them so small) that care should be taken not to overinterpret this finding.

We then allow for a linear spline, i.e. a different slope for the effect of genetic distance whether it is greater than the sample median of 0.104, or lower. Column (6) shows no evidence of such a differential effect (varying the spline threshold did not matter greatly). Finally, introducing a squared term in genetic distance (column 7) does not reveal much evidence of a nonlinear effect either. In sum, we find no evidence that the effect of genetic distance depends on other pair characteristics (such as geographic proximity) or that it is nonlinear.[44]

---

from the regression, although the (negative) effect of linguistic distance on conflict then becomes significant at the 10% level. These results are available upon request.

[44] In further tests available upon request, we allowed for nonlinear effects of geographic distance to capture the

Table 8 seeks to explain the intensity of militarized conflict as opposed to its incidence only. To do so, we modified the dependent variable in several ways. Column (1) simply uses the measure of the intensity of conflict from the Correlates of War dataset, rather than the binary transform of this variable we have been using so far. The dependent variable is the maximal intensity of conflict experienced by each pair over the 1816-2001 period (this variable can take any discrete value between 0 and 5). With least squares estimation, there is evidence that genetic distance bears a negative relationship with conflict intensity. However, column (2), limiting the sample to pairs having experienced conflict, demonstrates that genetic distance does not affect the intensity of conflict (among levels 3, 4 and 5) once we condition on the subsample with conflict: genetic distance works only on the extensive margin. This result rationalizes our focus on a binary measure of conflict rather than on the multinomial measure. In line with results in Table 3, instrumenting for current genetic distance using genetic distance based on the 1500 match of genetic groups to countries increases the estimated magnitude of the effect by over 60% (column 3).

In columns (4) and (5) we consider the determinants of war casualties, a variable obtained from the Correlates of War database and defined as the sum of all casualties experienced in a bilateral conflict over the 1816-2001 period. We find that genetic distance reduces total war casualties, but again this effect is almost entirely driven by the extensive margin, since genetic distance has a statistically insignificant effect on war casualties for observations with nonzero casualties. Finally, we consider the same dependent variable but instrumenting for genetic distance using the measure based on the 1500 match (column 6). Again, the effect increases, this time by about 50%, and it remains negative and highly significant statistically.

To summarize, the effect of genetic distance is very robust to using alternative measures of conflict. We also find no evidence that genetic distance affects the intensity of conflict conditional on a conflict occurring.

4.5 Panel Analysis for the 1816-2001 Period

For the remainder of this paper, we exploit the panel dimension of the data in order to control for time varying factors, such as whether a pair is composed of democratic countries, the intensity of

possibility that genetic distance may have captured the non-linear effect of physical distance, finding no evidence of this. We also allowed for an interaction term between genetic and geodesic distance, but this term was again found to be insignificant.

trade links between pairs, income differences, and other time-varying factors potentially affecting conflict. We begin by adding variables that are observed over the entire period (1816-2001) and in the next subsection turn to a more restricted set of variables available only after 1953.

The dependent variable is now C_{ijt}, an indicator of whether countries i and j were involved in an interstate conflict in year t. Table 4, column 1 presents baseline estimates from the panel specification of equation (20) over the period 1816-2001. This specification contains the same controls as the baseline cross-sectional regression, but is estimated over $517,251$ data points corresponding to multiple years observed on the underlying $13,175$ country pairs. Standard errors are now clustered at the country pair level. We replicate our basic findings, although the standardized magnitude of the effect of genetic distance falls by about half compared to the cross-sectional specification. A possible reason is that the share of observations where $C_{ijt} = 1$ in the panel sample is much smaller than the share of observations with $C_{ij} = 1$ in the cross-section. In the panel, there are $1,010$ wars and $3,728$ conflicts, or 0.20% and 0.72% of the sample, respectively, while in the cross-section 2.09% of the pairs were ever at war, and 5.65% of them were ever in conflict over 1816-2001.

In subsequent columns, we add time varying controls. In addition to indicators of conflict, the Correlates of War database includes several other useful time-varying bilateral variables such an indicator of whether a pair is linked by an active military alliance, the number of other wars occurring in a given year, and the number of peaceful years in a country pair (i, j) at each time t.[45] Including these variables does not greatly modify the effect of genetic distance (column 2), and the added controls bear coefficients with the expected signs.

In column 3, we include a dummy variable for whether both countries in the pair are democratic (defined as both having a Polity score in excess of 5). In this specification, the standardized magnitude of the effect of genetic distance is 8.52% of the mean percentage probability of conflict. The results provide evidence for a central tenet of liberal peace theory, namely the idea that democracies tend not to go to war with each other: the dummy for both countries being democratic has a negative and highly significant marginal effect, with roughly the same magnitude as that of genetic distance. The results also suggest that past findings on the pacifying effects of democracy did not capture the effects of cultural relatedness working through institutional similarity.

Finally, in column 4, we add the absolute difference in an index of national capabilities. For each country in a pair, the Correlates of War database provides an index of overall military capability

[45] These variables are also controls in Martin, Mayer and Thoenig (2008), Table 3, pp. 883.

(Singer, Bremer and Stuckey, 1972). This index is the average of underlying indices of iron and steel production, urban population, total population, total military expenditures, total military personnel and total energy production, and is meant to capture the country's ability to wage war successfully relative to other countries at a given point in time.[46] It is important for us to control for military capabilities because genetic distance may partly capture differences in capabilities. As discussed in the theoretical section, if similarity in military capabilities is persistent across generations, and if it affects the probability of conflict positively, omitting it may overstate the effect of genealogical relatedness on conflict working through more similar preferences over the object of dispute. We find in fact that including the absolute difference in the index of military capabilities barely reduces the standardized effect of genetic distance.[47]

4.6 Panel Analysis for the 1953-2000 Period

Several important correlates of war, such as measures of trade intensity and differences in income, are missing from our specification due to their lack of availability over the long time period covered by the baseline specification (1816-2001). In order to incorporate these additional controls, we focus on the 1953-2000 period for which various measures of trade and income are available. We continue to condition on the full set of controls from column 4 of Table 9 in all the regressions that follow.

A long tradition associated with liberal peace theory, going back to Montesquieu (1748) and Kant (1795), holds that extensive bilateral commercial links between countries reduce the probability of conflict, since valuable gains from trade would be lost in a militarized conflict. In an important contribution to the literature on interstate conflict, Martin, Mayer and Thoenig (2008, henceforth MMT) added an additional hypothesis: if the countries in a pair trade a lot with third parties, their bilateral trading link matters less, so controlling for bilateral trade, multilateral trade intensity should increase the probability of conflict among the countries in a pair. The issue we face is that the omission of these trade terms may bias the coefficient estimate on genetic distance,

[46] We also tried to examine what happens when we enter the absolute bilateral differences in the subcomponents of the index of national capabilities separately, or all of them jointly instead of the index. We found no systematic pattern in the effects of these various measures of differences in capabilities on the likelihood of conflict, and none of them - either separately or jointly - affected in any material way the coefficient on genetic distance.

[47] All of the robustness checks presented above in the cross-sectional context were carried out in the panel context as well, and were described in previous versions of this paper. They are available upon request.

to the extent that genetic distance and trade are correlated.

We obtained the data on bilateral and multilateral trade openness used in MMT's paper, and included their measures of trade in our baseline specification.[48] These measures include a metric of bilateral trade openness (the ratio of bilateral imports to GDP, averaged across the two countries in a pair), a metric of multilateral trade intensity (defined as the ratio of the sum of all bilateral imports from third countries to GDP, averaged between the two countries in a pair), and the interaction of each of these metrics with log geodesic distance. All of these measures were lagged by 4 years to limit the incidence of reverse causality running from conflict to trade, exactly as was done in MMT.

Results appear in Table 10. In column (1), we replicate the baseline specification for the smaller sample covering 1953-2000. The effect of genetic distance in this sample is slightly smaller than in the 1816-2001 sample: a standard deviation increase in genetic distance reduces the probability of conflict by 6.58% of this variable's mean. In column 2, we recover exactly the pattern of coefficients on the trade terms found in MMT: bilateral openness reduces conflict, multilateral openness raises conflict, and these effects are more pronounced quantitatively for pairs that are closer to each other. Thus, our findings lend further support to liberal peace theory, as recently amended by MMT. The effect of genetic distance falls a bit more in the specification that includes the trade terms, but it remains negative and highly significant statistically. In column (3), we include additional trade-related variables: a dummy for whether the two countries in a pair belong to a free trade area, and the number of GATT members in the pair. The coefficient on genetic distance is barely affected.

Another omitted variables concern stems from our results in Spolaore and Wacziarg (2009), where genetic distance was found to be robustly correlated with absolute differences in per capita income across pairs of countries. To the extent that differences in income capture power imbalances, or the extent of possible spoils of war, they may influence the probability of conflict (this could go in either direction: power imbalances may make a weaker prey easier to capture militarily, but also more willing to surrender peacefully). In column (5), we add the absolute value of the bilateral difference in log per capita income (the same variable used as a dependent variable in Spolaore and Wacziarg, 2009) to the specification that includes the broadest set of controls (including trade controls from MMT).[49] The coefficient on income differences is positive and significant, indicating

[48] The data was obtained from http://team.univ-paris1.fr/teamperso/mayer/data/data.htm

[49] The source for the income data is the Penn World Tables, version 6.1 (Heston, Summers and Aten, 2002).

that heterogeneity in income levels across the countries in a pair is conducive to conflict, but its inclusion does not affect the coefficient on genetic distance. Finally, column (6) substitutes the absolute difference in total GDP instead of differences in per capita GDP. This can be viewed as an alternative measure of differences in national capabilities. Heterogeneity in total GDP does not affect conflict propensity, and its inclusion does not affect the coefficient on genetic distance.

To summarize, the inclusion of a wide set of trade-related controls and of income differences, while confirming past results in MMT, does not change the basic message that relatedness has a positive effect on conflict.

5 Conclusion

The central insight of this paper is that populations that are more closely related are more likely to share preferences over rival goods, and are therefore more likely to enter into conflicts. Examining the empirical relationship between the occurrence of interstate conflicts and the degree of relatedness between countries, we found that populations that are genetically closer are indeed more prone to engage in militarized conflicts. This empirical result is large in magnitude and robust to controlling for several measures of geographic distance, income differences, and other factors affecting conflict, including measures of bilateral and multilateral trade and differences in democracy levels. These results provide strong evidence against the primordialist view that cultural and ethnic dissimilarity should breed war, plunder and a clash of civilizations. On the contrary, our findings suggest that war and peace across nations are (unhappy) family matters.

In theory, relatedness impacts conflict differently depending on the nature of the prize under dispute. An important distinction brought to the fore by this paper is that between private and public goods. Differences in preferences over private goods should lower the probability of conflict, while differences in preferences over public goods should increase it. Interstate conflict provides an ideal laboratory to test this idea, because most international conflicts are about rival goods: territory, resources, control of strategic routes. Thus, we expect to find that a broad measure of differences in intergenerationally transmitted traits - genetic distance - should be negatively correlated with the probability of conflict. Empirically, we do.

The fact that civil wars can span the spectrum of conflicts over rival to public goods might explain the mixed results found in the literature concerning the effect of ethnic divisions on the onset of civil wars. If all civil conflicts were about the type of public goods (location of the national capital, language of instruction in schools, etc.), we should expect similarity in preferences, and thus historical relatedness, to reduce conflict. Since civil wars are often also about private goods - such as territory or resource transfers from a central government - we expect ambiguous predictions on the effect of ethnic, cultural and historical relatedness on the onset of civil conflict. This ambiguous relationship is reflected in the empirical literature on this question, where the effect of ethnic divisions is not a settled matter. Further explorations of this interesting issue are left for future research.

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 (1997), "On the Number and Size of Nations," *Quarterly Journal of Economics*, 90(5), 1276-1296.

Alesina, A. and E. Spolaore (2005), "War, Peace and the Size of Countries", *Journal of Public Economics,* 89(7), 1333-1354.

Alesina, A. and E. Spolaore (2006), "Conflict, Defense Spending, and the Number of Nations", *European Economic Review* 50(1), 91-120.

Ashraf, Q. and O. Galor (2010), "The 'Out-of-Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *working paper*, Brown University, forthcoming in the *American Economic Review.*

Ashraf, Q. and O. Galor (2011), "Cultural Diversity, Geographical Isolation, and the Origin of the Wealth of Nations," *NBER Working Paper* #17640.

Barbieri, K. (2002), *The Liberal Illusion: Does Trade Promote Peace?*, University of Michigan Press, Ann Arbor.

Bisin, Alberto and Thierry Verdier (2000), "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits", *Quarterly Journal of Economics*, 105, 955-988.

Bisin, A. and T. Verdier (2001), "The Economics of Cultural Transmission and the Evolution of Preferences", *Journal of Economic Theory*, 97: 98-319.

Blainey, G. (1988), *The Causes of War*, Free Press, 3rd edition.

Boulding, K. E. (1962), *Conflict and Defense: A General Theory*, New York: Harper.

Boyd, R. and P. J. Richerson (1985), *Culture and the Evolutionary Process*, Chicago: University of Chicago Press.

Bowles, Samuel (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.

Bueno de Mesquita, B. and D. Lalman (1992), *War and Reason: Domestic and International Imperatives*. New Haven: Yale University Press.

Bueno de Mesquita, B., J. D. Morrow, R. M. Siverson, A. Smith (1999), "An Institutional Explanation of the Democratic Peace", *American Political Science Review*, 93 (4): 791-807.

Caselli, F., M. Morelli and D. Rohner (2012), "The Geography of Inter-State Resource Wars," *Working Paper*, London School of Economics,, April.

Cavalli-Sforza, L. L. and M. W. Feldman (1981), *Cultural Transmission and Evolution*, Princeton: Princeton University Press.

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.

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. and D. Ray (1994), "On the Measurement of Polarization," *Econometrica*, 62 (4): 819–51.

Esteban, J. and D. Ray (2011), "Linking Conflict to Inequality and Polarization," *American Economic Review*, 101 (4): 1345-74.

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. (1995), "Rationalist Explanations for War", *International Organization*, 49: 379-414.

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.

Ferguson, N. (2006), *The War of the World. Twentieth-Century Conflict and the Descent of the West*, New York: The Penguin Press.

Garfinkel, M. R. (1990), "Domestic Politics and International Conflict", *American Economic Review,* 84(5): 1294-1309.

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.

Gowa, J. (2000), *Ballots and Bullets: The Elusive Democratic Peace*, Princeton University Press, Princeton.

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.

Hess, G. D. and A. Orphanides (1995), "War Politics: an Economic Rational-Voter Framework", *American Economic Review* 85(4), 828-846.

Hess, G. D. and A. Orphanides (2001), "War and Democracy", *Journal of Political Economy*, 109: 776-810.

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.

Jackson M. and M. Morelli (2007), "Political Bias and War", *American Economic Review,* 97(4): 1353-1373.

Jones, D. M., S. A. Bremer, and J. D. Singer (1996), "Militarized Interstate Disputes, 1816–1992: Rationale, Coding Rules, and Empirical Patterns", *Conflict Management and Peace Science,* 15: 163-213.

Kant, I. (1795), *Toward Perpetual Peace*, New Haven: Yale University Press, 2006.

Kimura, M. (1968), "Evolutionary Rate at the Molecular Level", *Nature*, 217: 624-626.

Levy, G. and R. Razin (2004), "It Takes Two: An Explanation for the Democratic Peace", *Journal of the European Economic Association*, 2(1): 1-29.

Maoz, Z. and B. Russet (1993), "Normative and Structural Causes of the Democratic Peace, 1946-1986", *American Political Science Review*, 89.

Martin, P., T. Mayer and M. Thoening (2008) "Made Trade Not War?", *Review of Economic Studies*, 75: 865-900.

Maynes, C (1993), "Containing Ethnic Conflict", *Foreign Policy*, 90: 3-21.

Mecham, R. Q., J. Fearon and D. Laitin (2006), "Religious Classification and Data on Shares of Major World Religions", *unpublished*, Stanford University.

Michalopoulos, S. and E. Papaioannou (2010), "Divide and Rule or the Rule of the Divided? Evidence from Africa," CEPR Discussion Paper 8088.

Michalopoulos, S.and E. Papaioannou (2011), "The Long-Run Effects of the Scramble for Africa," CEPR Discussion Paper 8676.

Montalvo, J. G. and M. Reynal-Querol (2005), "Ethnic Polarization, Potential Conflict and Civil War", *American Economic Review*, 95(3): 796-816.

Montesquieu, C. de (1748), *The Spirit of the Laws,* Cambridge: Cambridge University Press, 1989.

Oneal, J. R. and B. M. Russett (1999a), "The Kantian Peace: The Pacific Benefits of Democracy, Interdependence, and International Organizations", *World Politics* 52(1): 1-37.

Oneal, J. R. and B. M. Russett (1999b), "Assessing the Liberal Peace with Alternative Specifications: Trade Still Reduces Conflict", *Journal of Peace Research*, 36(4): 423-442.

Oppenheimer, Stephen (2006), *The Origins of the British: A Genetic Detective Story*, New York: Carroll & Graf Publishers.

Polachek, S. (1980), "Conflict and Trade," *Journal of Conflict Resolution*, 24, 55-78.

Powell, R. (1999), *In the Shadow of Power: States and Strategies in International Politics.* Princeton: Princeton University Press.

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.

Russett, B.M. and J.R. Oneal (2001), *Triangulating Peace. Democracy, Interdependence, and International Organizations.* Norton: New York.

Schelling, T. (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.

Schneider, G, K. Barbieri and N.P. Gleditsch, eds. (2003), *Globalization and Armed Conflict*, Lanham: Rowman and Littlefield.

Singer, J.D. (1972), *The Wages of War, 1816-1965: A Statistical Handbook*, New York: Wiley.

Singer, J.D. (2000), "The Etiology of Interstate War: A Natural History Approach", in J.A. Vasquez, ed. (2000), *What Do We Know About War?* Lanham: Rowman & Littlefield.

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, 19-48.

Skaperdas, S. (1992), "Cooperation, Conflict and Power in the Absence of Property Rights", *American Economic Review* 82(4): 720-39.

Spolaore, E. (2008), "Civil Conflict and Secessions", *Economics of Governance*, 9(1): 45-63.

Spolaore, E. (2012), "National Borders, Conflict and Peace," chapter 30 in M. R. Garfinkel and S. Skaperdas (eds), *Oxford Handbook of the Economics of Peace and Conflict*, Oxford; Oxford University Press, 763-786.

Spolaore E. and R. Wacziarg (2009), "The Diffusion of Development", *Quarterly Journal of Economics*, 124(2): 469-529.

Spolaore E. and R. Wacziarg (2012), "How Deep are the Roots of Economic Development?", forthcoming, *Journal of Economic Literature.*

Stock, J. H. and F. Trebbi (2003), "Who Invented Instrumental Variable Regression?", *Journal of Economic Perspectives,* 17(3): 177-194.

Sumner, W. G. (1906), *Folkways*, New York: Ginn.

Tir, J. (2003), "Averting Armed International Con‡icts Through State-to-State Territorial Transfers", *Journal of Politics*, 65: 1235-1257.

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.

Vasquez, J. A. (1995), "Why Neighbors Fight? Proximity, Interaction or Territoriality", *Journal of Peace Research,* 32(93): 277-293.

Vasquez, J. A. ed. (2000), *What Do We Know About War?* Lanham: Rowman & Littlefield.

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.

Weede, E. (1973), "Nation-Environment Relations as Determinants of Hostilities Among Nations," *Peace Science Society (International) Papers*, 20: 67-90.

Wright, Q. (1942), *A Study of War* [1964 abridged ed.], Chicago: University of Chicago Press.

Wright, S. (1950), "The Genetical Structure of Populations", *Nature*, 166(4215): 247-249.

Yared, P. (2010), "A Dynamic Theory of War and Peace", *Journal of Economic Theory*, 145: pp. 1921-1950

Appendix

A.1. Peaceful Bargaining

In our basic model, the two states engage in conflict when both strongly care about the prize. However, conflict is costly, and both states would be better off if they could agree on an allocation of the prize that replicates the expected allocation from conflict, without bearing the actual costs from violent confrontation. For instance, if the prize is divisible, the two states would be better off by sharing it in proportion to their relative power - i.e., state i would obtain a share equal to π_i and state j would obtain a share equal to $(1 - \pi_i)$. If the prize is indivisible, the states could in principle agree to a lottery where each has a probability of winning the prize equal to its probability of winning the war, therefore saving the costs of going to war. However, even abstracting from issues of imperfect information, it might be extremely difficult to implement such a solution *ex post* (the loser may prefer to go to war after all). Even in the case of a divisible prize, states may have an incentive to unilaterally renege from the bargaining solution, and a war may occur as an equilibrium because each state would be better off fighting than surrendering when the other state fights. In fact, war may be the only equilibrium if each state faces a positive incentive to go to war unilaterally when the other state has agreed to a peaceful negotiation. In the absence of incentives to deviate unilaterally from peaceful bargaining, multiple equilibria may occur: war and peaceful bargaining.

In the latter case, more closely related populations, which share more recent common ancestors and hence may be more similar culturally, linguistically, etc., might be more successful at communicating and coordinating on the efficient equilibrium. If the probability of solving the conflict via peaceful bargaining is indeed higher for more closely related populations, this coordination effect could reduce or offset the main effect stemming from similarity in preferences. Then, the net effect of genetic distance on conflict would be ambiguous. However, coordinating on peaceful bargaining in an anarchic international environment, in the absence of credible commitment technologies, might be relatively rare. Moreover, the hypothesis that more closely related populations are better at coordination is purely speculative, and one could conceive of reasons why coordination may be harder among people who care more strongly about the same rival and excludable goods. Therefore, it is not clear, *ex ante*, whether such coordination effect would reduce or eliminate the main effect of relatedness highlighted in our model. As we have seen in Section 4, the empirical evidence is consistent with the main effect in our model dominating any countervailing effect from coordination on peaceful bargaining. These ideas are formalized below with a simple extension of our basic model of conflict over rival goods.

Consider an extension of the basic model, where peaceful bargaining can follow the choice of actions $\{C, F\}$ - which are now re-interpreted as {challenge, respond to challenge} rather than {challenge, fight}. Assume that, if state j challenges and state i responds to the challenge, each player can choose whether to bargain (B) or to go to war (W). If both choose "bargain," the prize is divided peacefully between the two states, and the two states obtain benefits equal to $\pi_i b_i$ and $(1-\pi_j)b_j$, respectively. That is because we assume that a state's bargaining power depends on its strength should negotiations break down (peaceful bargaining takes place "under the shadow of war").[50] If both states choose W, war follows, with the same payoffs as in the basic model. If state i chooses W while state j chooses B, war also follows, but with the following payoffs:

$$U_i\{W, B\} = (1+\xi)\pi_i b_i - c_i \tag{22}$$

and

$$U_j\{W, B\} = [1-(1+\xi)\pi_i]b_j - c_j \tag{23}$$

where

$$0 < \xi \leq \frac{1}{\pi_i} - 1 \tag{24}$$

The parameter ξ captures the increased probability of winning that results from being the initiator of the conflict, in the tradition of Schelling (1960).[51] By the same token, if state i chooses B in the sub-game, but state j chooses W, the payoffs are

$$U_i\{B, W\} = [1-(1+\xi)(1-\pi_i)]b_i - c_i \tag{25}$$

and

$$U_j\{B, W\} = (1+\xi)(1-\pi_i)b_j - c_j \tag{26}$$

Under these assumptions, if one state plays W, the other state is better off playing W rather than B, which implies that $\{W, W\}$ is a Nash equilibrium of the sub-game for all values of the

[50] This is a common assumption in the literature. For example, see Alesina and Spolaore (2005).

[51] Analogous results could be obtained by also assuming that the initiator of the conflict faces lower war costs. We abstract from this possibility to keep notation simple.

parameters. However, $\{W, W\}$ may or may not be the unique Nash equilibrium. If $\{W, W\}$ is the unique Nash equilibrium, the implications of this extension are the same as the basic model's. If $\{B, B\}$ is also a Nash equilibrium, war may be avoided if both states coordinate on the peaceful equilibrium. Therefore, our model is consistent with Fearon's (1995) discussion of war as emerging from an inability to commit to a Pareto-superior outcome. In our framework both states would be better off if each could commit to play B, but they can do that credibly only if $\{B, B\}$ is also a Nash equilibrium. For the symmetric case ($\pi_i = \frac{1}{2}$ and $c_i = c_j = c$), a necessary and sufficient condition for $\{B, B\}$ to be an equilibrium of the sub-game is

$$\xi \leq \frac{2c}{\min\{b_i, b_j\}} \tag{27}$$

The intuition for the above condition is straightforward: the parameter capturing the unilateral incentives to deviate from bargaining must be small enough for $\{B, B\}$ to be a Nash equilibrium of the sub-game. If $\{B, B\}$ is a Nash equilibrium of the sub-game, it is the unique coalition-proof Nash equilibrium. Three cases are possible: (i) states never coordinate on such an equilibrium even when the condition holds, (ii) states always coordinate on such equilibrium when available, and (iii) sometimes states coordinate, while other times they don't (coordination failure). Cases (i) and (ii) do not modify the implications of the basic model regarding the effect of relatedness on conflict. Relatedness is positively associated with the probability of war whenever conflict occurs - i.e., for all values of ξ in case (i), and for $\xi > \frac{2c}{\min\{b_i, b_j\}}$ in case (ii).

The effect of relatedness on conflict could in principle be modified in case (iii), *if* the likelihood of observing a coordination failure happened to depend on relatedness. For instance, coordination failure could be more likely across populations that are genealogically more distant, because their norms, habits, languages etc. would tend to be more different, and they might therefore find communication and coordination more difficult. If that were the case, such "coordination failure effect" would reduce the negative correlation between genetic distance and probability of conflict. However, *a priori*, and in the absence of a compelling theory of "equilibrium selection," there is no strong reason to expect that coordination failure would be less likely among more closely related populations. The relationship might even go in the opposite direction: coordination failure could be more likely between more closely related populations - for example, because of mistrust and animosity due to a history of previous conflicts over other rival goods. In the latter case, the effect of relatedness on conflict would be strengthened. As we have seen in the empirical section, the net effect of genetic distance on conflict is negative. This is consistent with two possibilities:

(a) coordination failure is not less likely among more closely related populations, (b) coordination failure is less likely among more closely related populations, but this effect is not large enough empirically to offset the main effect of relatedness on conflict highlighted by the basic model.

A.2. Conflict over Public Goods

In our basic model the prize is a rival and excludable good, and the type of the rival good is given (that is, it cannot be changed by either player). We now consider the polar case, where the prize is a pure public good, non-rival in consumption, and the player in control can choose whether the public good is of type A or B.[52] That is, the conflict is not about controlling access to the good (both players benefit from the good no matter who "owns" it), but about controlling the type (e.g., the characteristics of a public policy or service): the "winner" will select his/her favored type of public good. Hence, utilities from the public good are given as follows:

a) If player i and player j are of the same type, both obtain maximum benefits R from the good no matter who is in control:

$$b_i = b_j = R \tag{28}$$

b) If the two players are of different types, and player i is in control of the public good, the respective benefits are

$$b_i = R \tag{29}$$

$$b_j = (1 - |t_A - t_B|)R \tag{30}$$

c) conversely, if the two players are of different types, and player j is in control of the public good, we have:

$$b_i = (1 - |t_A - t_B|)R \tag{31}$$

$$b_j = R \tag{32}$$

Now, there is no reason for conflict between two players of the same type: if player i is of the same type as player j, player j will obtain the same utility as if he/she were in control of the good. In contrast, if player i is of a different type, player j could increase his/her utility by seizing control

[52] In this extension we refer to "players" rather than "states" consistently with our view that conflict over types of public goods is more likely to occur among agents engaged in intrastate conflict rather than interstate conflict.

of the good and changing the type. Hence, now a necessary condition for war is that the players are of *different* types. Then, player i's expected utility from going to war is

$$\pi_i R + (1 - \pi_i)[1 - |t_A - t_B|]R - c_i \tag{33}$$

and he/she will prefer to fight for

$$\pi_i R + (1 - \pi_i)(1 - |t_A - t_B|)R - c_i > (1 - |t_A - t_B|)R \tag{34}$$

which can be re-written as

$$\pi_i(|t_A - t_B|)R - c_i > 0 \tag{35}$$

By the same token, player j's will prefer war over "no challenge" for

$$(1 - \pi_i)(|t_A - t_B|)R - c_j > 0 \tag{36}$$

In the symmetric case ($\pi_i = 1/2$ and $c_i = c_j = c$) the two conditions become

$$c < (|t_A - t_B|)\frac{R}{2} \tag{37}$$

If the condition above is satisfied (i.e., the war costs are small enough), the probability that player i and player j engage in conflict is equal to the probability that they are not of the same type. For sibling populations ($d_g(i,j) = 1$), the probability that they are not of the same type is:

$$P\{i, j \mid d_g(i,j) = 1\} = 1 - F(\mu) = 1 - [\mu^2 + (1 - \mu)^2] \tag{38}$$

while for cousin populations ($d_g(i,j) = 2$), the probability that both are of different types is

$$P\{i, j \mid d_g(i,j) = 2\} = 1 - G(\mu) = 1 - [\mu^4 + 6\mu^2(1 - \mu)^2 + (1 - \mu)^4] \tag{39}$$

As we have shown in Section 2, $F(\mu) > G(\mu)$ for all $1/2 < \mu < 1$,which immediately implies

$$P\{i, j|d_g(i,j) = 1\} < P\{i, j \mid d_g(i,j) = 2\} \tag{40}$$

that is, we have:

Proposition 2A

When conflict is about the control of public-good types, the probability of violent conflict is higher between groups that are less closely related.

A.3 A General Framework

The two basic models (conflict over pure rival goods and conflict over pure public goods) can be viewed as two special cases of a more general framework where: (a) there may be externalities in consumption, and (b) the player in control of the prize may be able to change the good's type. Formally:

(a) when player i is in control of the prize, player j's benefits are δb_j and when player j is in control, player i's benefits are δb_i where $0 \leq \delta \leq 1$.

(b) when a player is in control of the prize of type A, he/she can change the type to B (and vice versa, can change type B to type A), with probability γ ($0 \leq \gamma \leq 1$).

Our basic model of conflict over rival goods is the case $\delta = \gamma = 0$, while the model of conflict over public goods in Appendix A.2 is the polar case $\delta = \gamma = 1$.

In general, two players with the same preferences will go to war with each other at low levels of δ (for all γ), while two players with different preferences will go to war with each other for high levels of γ, when $\delta > 0$. These results generalize the insights from the basic models: similarity in preferences leads to more conflict over goods with zero or low externalities (low δ), while dissimilarity in preferences leads to more conflict when agents in control of a non-rival good ($\delta > 0$) can change the good's type (high γ). Formally, we have:

Proposition A3.1

For all γ, there exists a critical $\delta^* = 1 - \frac{2c}{R}$ such that two players of the same type will go to war for $\delta < \delta^*$ and will not go to war for $\delta > \delta^*$.[53]

Proof. Two players of the same type $X (X = A, B)$ will not go to war over a good of type X if

$$\frac{1}{2}R + \frac{1}{2}\delta R - c < \delta R \tag{41}$$

and will not go to war over a good of type $Y \neq X$ if

$$\frac{1}{2}[\gamma R + (1-\gamma)(1-|t_A - t_B|)R] + \frac{1}{2}\delta[\gamma R + (1-\gamma)(1-|t_A - t_B|)R] - c < \delta[\gamma R + (1-\gamma)(1-|t_A - t_B|)R] \tag{42}$$

which can be re-written, respectively, as[54]

$$\delta < 1 - \frac{2c}{R} \tag{43}$$

[53] Multiple equilibria, with and without conflict, exist in the knife-edge case $\delta = \delta^*$.

[54] For $\delta = 0$, the condition below reduces to the condition for war in the case of pure rival goods: $c < \frac{1}{2}R$.

and

$$\delta < 1 - \frac{2c}{R[1-(1-\gamma)|t_A - t_B|]} \tag{44}$$

For all $0 \leq \gamma \leq 1$, we have

$$1 - \frac{2c}{R} \geq 1 - \frac{2c}{R[1-(1-\gamma)|t_A - t_B|]} \tag{45}$$

Therefore, for all $\delta > \delta^* \equiv 1 - \frac{2c}{R}$ we also have $\delta > 1 - \frac{2c}{R[1-(1-\gamma)|t_A - t_B|]}$, and no war ever takes place between two players with the same preferences. In contrast, for $\delta < \delta^*$, the two players will go to war. QED.

In contrast, conflict between players with different preferences is characterized by the following proposition:

Proposition A3.2

For all $\delta > 0$, two players with different preferences will go to war for $\gamma > \gamma^*$, where[55]

$$\gamma^* = \frac{1}{|t_A - t_B|} \min\{1 - \frac{1}{\delta}(1 - \frac{2c}{R}) \; ; \frac{2c}{R} - (1-\delta)[1 - |t_A - t_B|]\} \tag{46}$$

Proof. When two players have different preferences, the player whose preferred type is the same as the prize will go to war if

$$\frac{1}{2}R + \frac{1}{2}\delta[\gamma(1 - |t_A - t_B|)R + (1-\gamma)R] - c > \delta[\gamma(1 - |t_A - t_B|)R + (1-\gamma)R] \tag{47}$$

while the other player will go to war if

$$\frac{1}{2}[\gamma R + (1-\gamma)(1 - |t_A - t_B|)R] + \frac{1}{2}\delta(1 - |t_A - t_B|)R - c > \delta(1 - |t_A - t_B|)R \tag{48}$$

The above equations can be re-written as

$$\gamma > \frac{1 - \frac{1}{\delta}(1 - \frac{2c}{R})}{|t_A - t_B|} \tag{49}$$

and

$$\gamma > \frac{\frac{2c}{R} - (1-\delta)(1 - |t_A - t_B|)}{|t_A - t_B|} \tag{50}$$

Both conditions hold if $\gamma > \gamma^*$. QED.

---

[55] In the case $\delta = \gamma = 1$, the condition below reduces to the condition for war in the case of pure public goods: $c < (|t_A - t_B|)\frac{R}{2}$

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)
	Conflict, univariate specification	**Conflict, baseline specification**	**Conflict, baseline specification IV**	**War, baseline specification**	**War, baseline specification IV**
Fst genetic distance, weighted	**-57.3760** (-17.800)**	**-19.8786** (-9.317)**	**-30.6802** (-8.843)**	**-6.3389** (-7.478)**	**-8.6043** (-5.746)**
Log geodesic distance		-1.6281** (-5.567)	-1.0182** (-3.090)	-0.2929* (-2.505)	-0.1728 (-1.349)
Log absolute difference in longitudes		0.1424 (0.731)	-0.0677 (-0.336)	-0.0197 (-0.254)	-0.0629 (-0.787)
Log absolute difference in latitudes		-0.1130 (-0.887)	-0.1312 (-1.002)	-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)	-0.6406** (-5.531)	-0.6500** (-5.635)
Number of island countries in the pair		0.8212** (2.923)	0.8762** (3.005)	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)	-0.0154 (-0.128)	-0.0199 (-0.161)
Log product of land areas in square km		0.8940** (18.992)	0.9045** (17.145)	0.3132** (17.452)	0.3201** (9.755)
1 for pairs ever in colonial relationship		7.3215** (5.094)	7.6147** (3.175)	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)	1.0952* (2.424)	1.1373 (1.564)
# of observations	13,175	13,175	13,175	13,175	13,175
Pseudo-R^2	0.075	0.275	-	0.236	-
Standardized effect (%)	-68.81	-23.84	-36.79	-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 – Sample breakdown by region
(Dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline specification	**Europe, with Europe Fst Gen. Dist.**	**Removing all European countries**	**Asia**	**Africa**	**America**
Fst genetic distance, weighted [a]	-19.8786 (9.317)**	-1,494.7018 (2.711)**	-12.9864 (6.892)**	-39.4754 (1.465)	-7.3404 (0.458)	-49.8330 (2.339)*
Log geodesic distance	-1.6281 (5.567)**	1.4474 (0.279)	-0.7379 (2.845)**	-1.8233 (0.939)	-2.0350 (1.829)	-1.9795 (0.824)
1 for contiguity	15.4610 (10.095)**	9.7531 (1.021)	15.7907 (9.693)**	20.8535 (3.867)**	18.1649 (4.732)**	41.2895 (4.173)**
# of observations	13,175	291	7,777	860	848	581
Pseudo-R^2	0.275	0.428	0.350	0.392	0.354	0.382
Standardized effect (%)	-23.839	-38.396	-19.706	-12.852	-3.684	-21.472

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.
[a] : Weighted genetic distance in all columns except column (2), where Fst genetic distance between plurality groups from the European genetic distance matrix is entered instead. The coefficient is larger because Fst genetic distance within Europe has a much smaller range than in the World matrix (the standardized magnitude is also larger but in a comparable order of magnitude as for the rest of the World).
Controls: In addition to reported coefficients, every column includes 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=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 5 – Sample breakdown by historical sub-period
(Dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(4)	(5)	(6)
	1816-2001 baseline	**1816-1900**	**1901-2001**	**1946-2001**	**1919-1989**	**1990-2001**
Fst genetic distance, weighted	**-19.8786 (9.317)****	**-1.1213 (5.750)****	**-18.8533 (8.879)****	**-11.8279 (-6.933)****	**-13.5510 (7.537)****	**-3.9197 (-5.027)****
Log geodesic distance	-1.6281 (5.567)**	-0.0682 (2.418)*	-1.6896 (5.848)**	-1.0813 (-5.185)**	-1.0343 (4.229)**	-0.3474 (-4.706)**
1 for contiguity	15.4610 (10.095)**	0.3185 (2.849)**	15.0125 (10.047)**	9.9743 (9.216)**	10.1606 (8.636) **	3.3041 (7.686)**
Pseudo-R^2	0.275	0.286	0.271	0.280	0.252	0.331
Standardized effect (%)	-23.839	-9.439	-23.331	-19.69	-20.668	-16.04

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 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.
13,175 observations used in all columns.
Controls: In addition to reported coefficients, every column includes 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=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 6 – 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 7 - Nonlinearities and sample splits
(Dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	**Excluding contiguous pairs**	**Major power interaction**	**Proximity interaction**	**Contiguity interaction**	**Spline**	**Quadratic**
Fst genetic distance, weighted	**-19.8786 (9.317)****	**-18.5357 (9.379)****	**-20.4475 (9.274)****	**-20.5701 (9.270)****	**-20.4463 (9.463)****	**-17.3704** (-3.904)**	**-18.3955 (3.093)****
Fst Gen. Dist * major power dummy			-3.1786 (0.517)				
Dummy=1 if at least one country is a major power			4.2005 (5.875)**				
Fst Gen. Dist. * proximity				7.8304 (1.689)			
Fst Gen. Dist. * contiguity					30.8443 (2.432)*		
Fst Gen. Dist * dummy for FST GD > median						-2.1460 (-0.637)	
Squared Fst genetic distance, weighted							-6.7332 (0.258)
Log geodesic distance	-1.6281 (5.567)**	-1.4809 (5.065)**	-1.3552 (4.746)**	-1.4900 (4.982)**	-1.6451 (5.642)**	-1.6317** (-5.580)	-1.6325 (5.531)**
1 for contiguity	15.4610 (10.095)**		15.6847 (10.214)**	15.2255 (10.056)**	10.1116 (5.971)**	15.4360** (10.102)	15.4319 (10.095)**
Observations	13,175	12,928	13,175	13,175	13,175	13,175	13,175
Pseudo-R^2	0.275	0.202	0.287	0.275	0.276	0.275	0.275
Standardized effect (%)	-23.839	-27.343	-24.521	-24.668	-24.520	-20.83	-22.060

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.
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.
Controls: In addition to reported coefficients, every column includes 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=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 8 – Regressions explaining the intensity of conflict
(Dependent variable and estimator as described in the second row)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS on maximal conflict intensity	**Same as (1) for the subsample with conflict**	**Maximal conflict intensity, IV with 1500 GD**	**OLS on index of casualties**	**Same as (4) for the subsample with casualties index > 0**	**Total casualties index, IV with 1500 GD**
Fst genetic distance, weighted	**-1.1810** **(9.340)****	**0.4048** **(0.801)**	**-1.9351** **(9.977)****	**-2.8008** **(5.397)****	**15.7310** **(0.942)**	**-4.5377** **(5.097)****
Log geodesic distance	-0.1501 (5.103)**	-0.0296 (0.530)	-0.0959 (2.984)**	-0.4262 (2.195)*	-2.3870 (0.982)	-0.3016 (1.424)
1 for contiguity	1.9217 (13.879)**	-0.1019 (1.121)	1.9290 (13.911)**	5.3173 (5.199)**	-3.8857 (1.170)	5.3342 (5.217)**
Constant	0.1134 (0.546)	3.2893 (9.463)**	-0.1842 (0.821)	-0.7472 (0.589)	-23.8304 (1.655)	-1.4326 (1.029)
Observations	13,175	756	13,175	13,175	406	13,175
Adjusted R^2	0.173	0.046	0.171	0.064	0.131	0.064
Beta coefficient on Fst GD	-8.011	3.132	-13.126	-4.223	4.346	-6.842

Robust t statistics in parentheses; * significant at 5%; ** significant at 1%.
The standardized beta is the effect of a standard deviation change in genetic distance as a percentage of the standard deviation of the dependent variable).
Controls: In addition to reported coefficients, every column includes 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=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 9: Panel analysis, 1816-2001
(Dependent variable: dichotomous indicator of conflict)

	(1)	(2)	(3)	(4)
	Baseline specification	**Add some time-varying controls**	**Add dummy for both democracies**	**Add difference in national capabilities index**
Fst genetic distance	-1.3230** (-5.796)	-0.9305** (-8.642)	-0.9313** (-8.922)	-0.8092** (-8.417)
Log geodesic distance	-0.1518** (-4.671)	-0.0743** (-4.379)	-0.0735** (-4.487)	-0.0534** (-3.998)
Log absolute difference in longitudes	-0.0165 (-0.796)	-0.0027 (-0.270)	-0.0003 (-0.029)	-0.0163 (-1.933)
Log absolute difference in latitudes	-0.0607** (-3.309)	-0.0280** (-3.100)	-0.0250** (-2.927)	-0.0258** (-3.474)
1 for contiguity	0.8463** (7.235)	0.4443** (7.760)	0.4227** (7.760)	0.4862** (8.399)
Number of landlocked countries in the pair	-0.2059** (-6.224)	-0.1267** (-7.541)	-0.1197** (-7.553)	-0.1009** (-6.800)
Number of island countries in the pair	0.1720** (4.371)	0.0503** (2.593)	0.0551** (2.969)	0.0540** (3.116)
1 if pair shares at least one sea or ocean	0.0674 (1.648)	0.1002** (4.212)	0.1029** (4.501)	0.0679** (3.364)
Log product of land areas in square km	0.0979** (13.164)	0.0544** (15.532)	0.0511** (15.762)	0.0376** (12.083)
1 for pairs ever in colonial relationship	0.2483** (3.066)	0.1152** (2.797)	0.1478** (3.413)	0.1270** (3.227)
1 if countries were or are the same country	0.0229 (0.262)	0.0457 (1.005)	0.0444 (1.021)	0.0679 (1.592)
Number of peaceful years		-0.0070** (-14.021)	-0.0066** (-13.545)	-0.0059** (-13.687)
Number of other conflicts in year t		0.0037** (16.334)	0.0035** (16.748)	0.0035** (18.425)
Dummy for alliance active in year t		-0.0667** (-5.150)	-0.0593** (-4.686)	-0.0604** (-5.505)
1 if both countries are democracies (polity2>5)			-0.0935** (-8.670)	-0.0910** (-8.088)
Absolute difference in National Capabilities Index				1.1408** (13.621)
Pseudo-R^2	0.210	0.295	0.300	0.321
Standardized effect	-12.11	-8.513	-8.521	-7.395

Robust t statistics in parentheses (clustering at the country pair level); * 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. Probit marginal effects reported throughout. For dummy variables, marginal effects are for discrete changes from 0 to 1. All marginal effects were multiplied by 100 for readability (underlying average probability of conflict is 0.72%).
All specifications were estimated with 517,251 observations from 13,175 country pairs.

Table 10: Post-1953 analysis, controlling for trade variables and absolute income differences (Dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(5)	(6)
	Baseline with common sample	**Control for trade terms**	**Additional trade terms**	**Add per capita income difference**	**Add total income difference**
Fst genetic distance	**-0.5195** (-6.964)**	**-0.3419** (-5.784)**	**-0.3253** (-5.475)**	**-0.3689** (-5.335)**	**-0.3301** (-4.512)**
Log geodesic distance	-0.0120 (-1.793)	0.0056 (0.795)	0.0052 (0.721)	0.0015 (0.223)	0.0031 (0.434)
1 for contiguity	0.2856** (6.907)	0.2396** (7.494)	0.2325** (7.469)	0.2932** (7.620)	0.2527** (7.235)
1 if both countries are democracies	-0.0412** (-5.140)	-0.0321** (-5.257)	-0.0249** (-3.845)	-0.0214** (-3.073)	-0.0257** (-3.699)
Absolute difference in national capabilities index	0.4063** (5.018)	0.1641* (2.116)	0.1516* (1.983)	0.1758* (1.968)	0.2291* (2.325)
Log bilateral openness, t-4		-0.0405** (-4.521)	-0.0386** (-4.283)	-0.0349** (-3.844)	-0.0387** (-3.996)
Log multilateral openness, t-4		0.0459 (1.523)	0.0512 (1.735)	0.0010 (0.033)	0.0142 (0.423)
Log distance * log mult. openness		-0.0079* (-2.034)	-0.0085* (-2.249)	-0.0027 (-0.681)	-0.0045 (-1.062)
Log distance * log bilateral openness		0.0052** (4.650)	0.0050** (4.481)	0.0045** (3.910)	0.0050** (4.076)
Dummy for zero trade, t-4		-0.0176* (-2.524)	-0.0173* (-2.557)	-0.0146 (-1.947)	-0.0158* (-2.050)
Free trade area (full set)			-0.0242** (-2.824)	-0.0217* (-2.260)	-0.0230* (-2.533)
# of GATT members			-0.0150** (-3.723)	-0.0166** (-3.893)	-0.0177** (-4.149)
Absolute diff. in log p.c. income				1.6054** (4.650)	
Absolute diff. in total GDP					-0.0519 (-0.195)
# of observations	226,357	226,357	226,357	202,523	202,523
(# of pairs)	(9,127)	(9,127)	(9,127)	(9,127)	(9,127)
Pseudo R^2	0.341	0.351	0.354	0.357	0.352
Standardized effect	-6.576	-4.328	-4.118	-5.248	-4.695

Robust t statistics in parentheses (clustering at the country pair level); * 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. 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.
Controls: In addition to reported coefficients, every column includes 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=1 if pair shares at least one sea or ocean, log product of land areas in square km, dummy=1 if both countries are democracies (polity2>5), dummy=1 for pairs ever in colonial relationship, dummy=1 if countries were or are the same country, number of peaceful years, number of other wars in year t, dummy for alliance active in year t.

Figure I - Population Tree

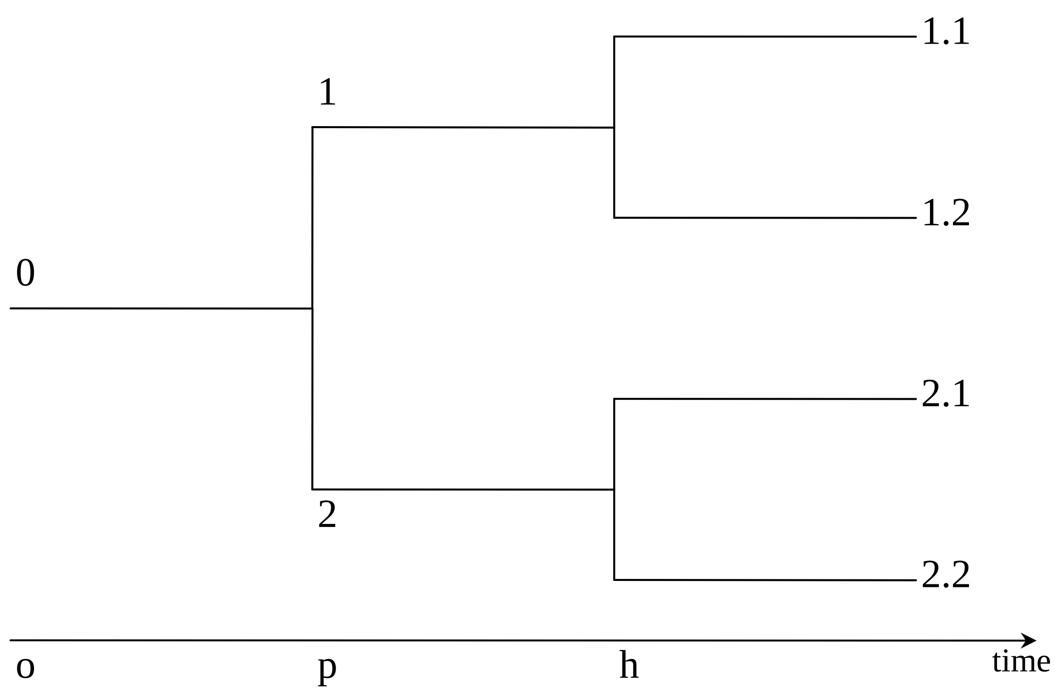
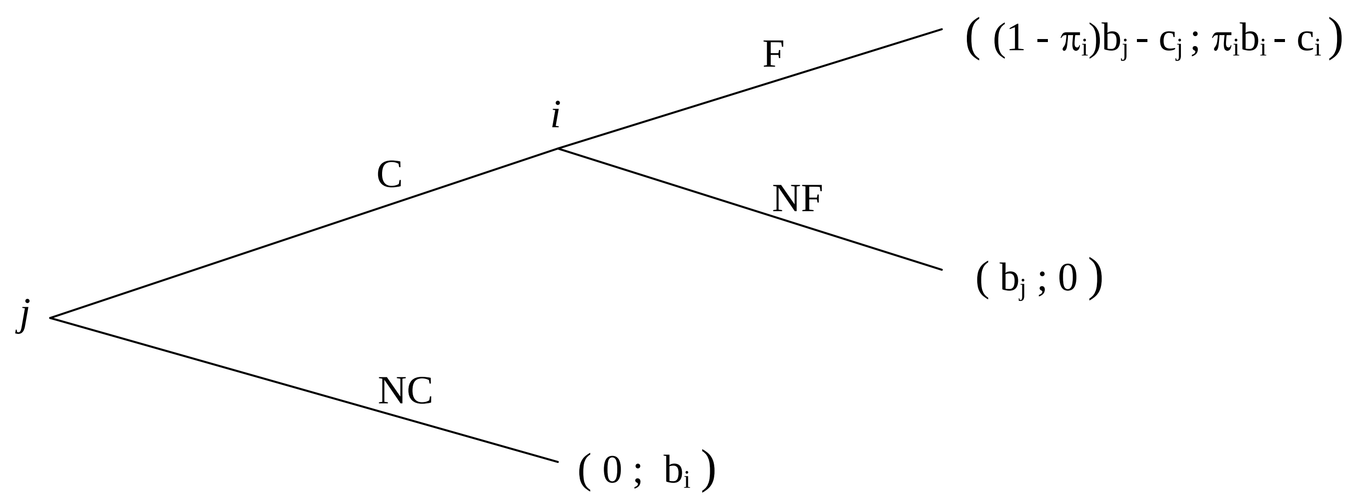


Figure II – Extensive-form Game



In parentheses: (state *j*'s payoff ; state *i*'s payoff)

Figure III – An Example