War and Relatedness*

Enrico Spolaore
Tufts University, NBER and CESIfo

Romain Wacziarg
UCLA, NBER and CEPR

April 2010

Abstract

We examine the empirical relationship between the occurrence of interstate conflicts and the degree of relatedness between countries, showing that genetically closer populations are more prone to go to war with each other, even after controlling for a wide set of measures of geographic distance and other factors that affect conflict, including trade and democracy. We provide a theoretical framework consistent with these findings. In our model, genealogical relatedness between populations has a positive effect on their conflict propensities because more closely related populations, on average, share a wider set of common issues over which disputes may emerge.

---

*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, Michelle Garfinkel, Tim Groseclose, Yannis Ioannides, Michael Klein, Philippe Martin, Deborah Menegotto, Massimo Morelli, Daniele Paserman, Ken Shepsle, Vickie Sullivan and seminar participants at numerous universities and 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.

1 Introduction

Militarized conflicts have been among the most dramatic and costly events in human history, and at the center of an enormous historical and political literature for centuries.[1] In recent years, economists and political scientists have started to use formal theoretical tools and systematic empirical analyses to provide insights into the determinants of conflicts and wars.[2] Great progress has been made in our understanding of the effects of economic and political factors - such as trade and democracy - on the likelihood of international conflict.[3] Nonetheless, wars continue to be elusive phenomena, and fundamental questions about their roots remain open. A key question, which has not yet received a satisfactory empirical answer, is whether armed conflicts are more or less likely to emerge between populations that differ along cultural and historical dimensions, such as ethnicity, language, and religion.

In this paper we present a new theoretical approach, new data and new empirical findings shedding light on the determinants of international conflict. We use information about human genetic distance - a summary statistic of very long-run historical relatedness between populations - to explore the relationship between kinship and conflict.[4] Genetic distance measures the difference

---

[1] For recent examples, see Blainey (1988), Keegan (1984), Ferguson (2006) and Nye (2008).

[2] Classic contributions are Schelling (1960) and Boulding (1962). More recent economic formalizations of conflict and wars include, for example, Garfinkel (1990), Hirshleifer (1991, 1995), Skaperdas (1992), Hess and Orphanides (1995, 2001), Jackson and Morelli (2009). Garfinkel and Skaperdas (2006) provide an overview of the economics literature on conflict. Influential contributions by political scientists on the formal theory of conflict include Bueno de Mesquita and Lalman (1992), Fearon (1995) and Powell (1999). Systematic empirical work on interstate conflict was pioneerd by Wright (1942), Richardson (1960) and Singer (1972). For discussions of the recent empirical literature on the correlates of war see Vasquez (2000) and Schneider, Barbieri and Gleditsch (2003).

[3] The liberal peace view that trade and democracy should reduce the risk of war goes back to Montesquieu (1748) and Kant (1795), and has been the subject of a vast literature (e.g., see Oneal and Russett, 1999a and Russett and Oneal, 2001). Contributions on the empirics of trade and conflict include Polacheck (1980), Oneal and Russett (1999b), Barbieri (2002), and Martin, Mayer and Thoenig (2008), among many others. On the democratic peace hypothesis see, for example, Maoz and Russett (1993), Bueno de Mesquita et al. (1999), Gowa (2000), and Levy and Razin (2004).

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

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, as in 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. In other words, divergence in neutral genes provides information about *lines of descent*: genetic distance is a summary measure of general relatedness between populations. Heuristically, the concept is analogous to relatedness between individuals: two siblings are more closely related than two cousins because they share more recent common ancestors - their parents rather than their grandparents. Since 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 result is that, surprisingly, genetic distance significantly reduces the risk of conflict, and this 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 a wide range of geographic measures, 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 having been in conflict between 1816 and 2000 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 1500, to account for measurement error and possible endogeneity issues due to post-1500 migrations and population mixing. The reduction in the probability of a pair being in conflict *in any given year*, as a result of a one standard deviation increase in genetic distance, is equal to 8.52% of the mean percentage probability of conflict (without instrumenting) and 13.19% (with IV).

These findings are consistent with a simple theoretical framework in which the degree of genealogical relatedness between populations has a positive effect on their conflict propensities, because closely related populations, on average, tend to share common traits and preferences, to interact with each other more, and to care about a larger set of common issues. In principle, such a

---

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 variable regression (and Sewall might have contributed to that discovery as well).

conflict-generating effect could be offset by countervailing forces. More closely related populations could also have closer ideal points or could be better at coordinating on peaceful equilibria. However, in the data these other forces, if they exist, do not seem to be strong enough to counteract the main effect stemming from the greater set of common issues arising among genetically related populations. In a nutshell, from a long-term world-wide perspective, issues of war and peace are (unhappy) family matters.[5]

This paper builds on a large and diverse literature. Broad questions about cultural distance, relatedness and conflict are probably as old as wars themselves, but have received increasing attention following the recent debate over the clash of civilizations (Huntington, 1993) and surging concerns about ethnic conflict within and across countries. For instance, Maynes (1993, p. 5) writes: "Animosity among ethnic groups is beginning to rival the spread of nuclear weapons as the most serious threat to peace that the world faces". Several commentators have wondered whether there may be a general tendency towards violent confrontation between populations that are ethnically distant. For example, Bremer (2000, p. 27), referring to evidence from social psychology, wonders 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 related would tend to fight less with each other. In contrast, others have emphasized instrumentalist views of ethnicity, implying that such differences should not be closely correlated with inter-group conflict (e.g., Merton, 1957). A related hypothesis, proposed but not tested by Gleditsch and Singer (1975), is that the paramount force in conflict is geographical contiguity, and that, controlling for contiguity, one would not find a significant correlation between cultural relatedness and interstate conflict (see Henderson, 1997, for a review of this debate). At the same time, the few scholars who have attempted to estimate the effects of common culture, language or religion on international conflict have found little or no evidence that such variables are systematically associated with a lower probability of conflict.[6] In their influential study on conflict within states, Fearon and Laitin

---

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

[6] For example, see Richardson (1960, p. 296), who found no general pacifying effect for either common language or common religion, and Henderson (1997), who, controlling only for contiguity, found a negative association between a measure of religious similarity and interstate conflict, and a positive (but insignificant) correlation between a measure of ethnic similarity and conflict. See also the more recent contribution of Gartzke and Gleditsch (2006).

(2003) also found no evidence that ethnically diverse states would be more likely to experience civil conflict.

Our results go further in casting doubts over primordialist theories, as we show not only that their predictions are falsified when applied to interstate conflict, but that the effect goes into the opposite direction. The negative effect of genetic distance holds when controlling for a vast range of geographic measures (contiguity, geodesic distance, latitudinal and longitudinal differences, and other measures of geographic barriers), contrary to Gleditsch and Singer's (1975) hypothesis that geographic proximity should be the predominant force in international conflict. It seems that the paramount effect attributed by some scholars to geographic proximity may in part be due to its correlation with historical relatedness. Once genetic distance is taken into account, geographic variables have smaller effects (although they remain significant). As already mentioned, the effect of genetic distance is higher - and the effects of geography smaller - when instrumenting for modern genetic distance using historical (1500 C.E.) genetic distance. We also find that the effect of genetic distance is larger for non-territorial conflicts relative to territorial conflicts, which further suggests an independent role for historical relatedness, in addition to purely geographic considerations. The effect of genetic distance is robust to controlling for other measures of similarity, such as religious and linguistic distance, and for differences in income per capita across countries. Interestingly, religious distance also reduces the likelihood of conflict. This would be hard to rationalize within a clash-of-civilizations view, but is consistent with the predictions of our common-issues model.

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. 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. We are therefore able to address one of the most important criticisms of the empirical work 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 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 historical 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 and the likelihood of international conflict and wars.[7] 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 diffusion of economic development from the world technological frontier. Desmet et al. (2007) 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.[8] More broadly, our paper is related to the evolutionary literature on cultural transmission of traits and preferences (e.g., Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985; Richerson and Boyd, 2004; for economic analyses of cultural transmission, see for instance Bisin and Verdier, 2000, 2001).

The rest of the paper is organized as follows. Section 2 presents a stylized model of conflict and relatedness (extensions are included in Appendix 1). Section 3 introduces the data and methodology. Section 4 discusses the empirical findings. Section 5 concludes.

---

[7] In general, 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.

[8] Desmet et al. (2007) 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.

2 A Model of Conflict and Relatedness

In this section we provide an analytical framework linking international disputes, the probability of armed conflict, and long-term relatedness between populations. We model three effects: a common-issues effect, a disagreement effect, and a coordination effect. The common-issues effect unambiguously implies a positive relationship between conflict and relatedness: populations that are more closely related are more likely to share similar tastes and preferences, hence to care about a larger set of common issues, and - *ceteris paribus* - to engage in military disputes over those issues. In contrast, as detailed in Appendix 1, the disagreement effect (whereby genealogically distant populations may disagree more over a given set of issues) and the coordination effect (whereby genealogically distant populations may find it harder to coordinate on a peaceful equilibrium) may go in the opposite direction. Empirically, we find that more closely related populations do in fact engage in more conflict with each other, which is consistent with the simplest version of our model focusing on the common-issues effect. Therefore, our theory provides a simple economic interpretation for the surprising fact that more closely related populations fight more with each other, even when controlling for other possible determinants of conflict.

2.1 Preferences and Common Issues

Consider two sovereign states (1 and 2), facing a set of issues M.[9] Each issue $k \in M$ can takes value $x(k) \in X(k)$, where M and $X(k)$ are sets of real numbers. Each state i's utility function is:

$$U_i = -\int_{k \in M} \alpha_i(k)|x(k) - x_i^*(k)|dk - c_i \tag{1}$$

where $x(k)$ is the actual outcome for issue k, $x_i^*(k)$ is state i's most preferred outcome, $\alpha_i(k) \geq 0$ is the weight that state i attributes to issue k, and c_i denotes net costs from conflict (which are zero if disputes are solved peacefully, positive otherwise). We introduce the following definitions:

Definition 1: Issue k is a *common issue* between the two states if both states care about issue k - that is, if and only if $\alpha_1(k) > 0$ and $\alpha_2(k) > 0$.

Definition 2: A common issue k is *disputed* when the two states prefer different outcomes: $x_1^*(k) \neq x_2^*(k)$, where $\Delta(k) \equiv |x_1^*(k) - x_2^*(k)|$ is the difference between ideal outcomes. Two states face a dispute when one or more common issues are disputed.

---

[9] For simplicity we treat a state - or, equivalently, its government - as a unitary agent.

2.2 The Resolution of Disputes

Disputes between the two states are resolved either peacefully or violently. When a dispute is resolved peacefully, neither state bears conflict costs ($c_i = 0$, $i = 1, 2$), and the outcome for each disputed issue k is:

$$x(k) = \beta x_1^*(k) + (1 - \beta) x_2^*(k) \tag{2}$$

where β denotes state 1's bargaining power in a peaceful dispute ($0 \leq \beta \leq 1$). If a violent conflict occurs, all disputed issues are resolved according to the winner's preferences. Let P denote the probability that state 1 will win in a violent conflict and set $x(k) = x_1^*(k)$ for all $k \in M$, while state 2 will win and set $x(k) = x_2^*(k)$ with probability $1 - P$.[10] Then the expected outcome of a violent dispute for each k is:

$$x(k) = P x_1^*(k) + (1 - P) x_2^*(k) \tag{3}$$

Each state can choose whether to "start a conflict" (strategy C) or "not to start a conflict" (strategy N). Peace results if and only if both states choose N, in which case all issues are settled peacefully. If both states choose C, $P = \pi$, where $0 \leq \pi \leq 1$, $c_i = c_i(C, C) > 0$. If state 1 chooses C while state 2 chooses N, we assume $P = \pi + \sigma_1$ where $0 < \sigma_1 \leq (1 - \pi)$. The parameter σ_1 captures the increased probability of winning that results from being the initiator of the conflict, in the tradition of Schelling (1960). The costs of conflict are then $c_1(C, N) \geq 0$ and $c_2(C, N) \geq c_2(C, C)$.[11] Analogous assumptions hold when state 1 chooses N but state 2 chooses C.[12] The game in normal form is summarized in Figure 1.

Under these assumptions, if one state plays C, the other state is better off playing C rather than N, which implies that (C, C) is a Nash equilibrium for all values of the parameters. However, (C, C) may or may not be the unique Nash equilibrium. If (C, C) is the unique Nash equilibrium, war occurs with certainty. If (N, N) 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

---

[10] Since utility functions are linear, we will not distinguish between ex-ante (expected) outcomes and ex-post (actual) outcomes in our notation, and denote both with $x(k)$.

[11] This assumption means, quite reasonably, that when state 2 enters into a conflict unwillingly, it will face conflict costs at least as high as if it had decided to start the conflict willingly (i.e., if it had selected C rather than N).

[12] When state 1 chooses N and state 2 chooses C, $P = \pi - \sigma_2$, with $0 < \sigma_2 \leq \pi$, $c_1 = c_1(N, C) \geq c_1(C, C)$ and $c_2 = c_2(N, C) \geq 0$.

framework both states would be better off if each could commit to play N, but they can do that credibly only if (N, N) is also a Nash equilibrium. To simplify the analysis, we assume:

(i) Symmetry ($\sigma_1 = \sigma_2 = \sigma$ and $c_1(C, N) = c_2(N, C) = \phi$).

(ii) Peaceful bargaining "under the shadow of war," (that is, a state's bargaining power depends on its strength should negotiations break down), which implies $\beta = \pi$.[13]

We can then define $\dfrac{\phi}{\sigma} \equiv \omega$. The parameter ω captures the relative cost of starting a war, increasing in the cost of going to war (ϕ) and decreasing in the temptation to start a war (σ). Then, it is immediate to show that:[14]

Remark 1: *The peaceful outcome* (N, N) *is a Nash equilibrium if and only if:*

$$\max_{i=1,2}\{\int_{k\in M} \alpha_i(k)\Delta(k)dk - \omega\} \leq 0 \tag{4}$$

In contrast, if $\max_i\{\int_{k\in M} \alpha_i(k)\Delta(k)dk - \omega\} > 0$*, conflict* (C, C) *is the unique Nash equilibrium.*

For a given relative cost of starting a war (measured by ω), violent conflict is more likely to be the unique Nash equilibrium the larger the set of common issues under dispute, and the larger the extent the two states care about those issues. What is the probability of an actual conflict between states, and how does it depend on long-term relatedness? We address these questions in the rest of this section.

2.3 Relatedness and the Probability of Conflict

So far the set of common issues under dispute has been taken as given. We will now consider the link between common issues and long-term connections between populations. The general idea is that if preferences over issues are persistent across time, and current populations inherit such tastes with variation from their ancestors, on average populations that are more closely related will be more likely to share a larger range of common issues.

---

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

[14] This is a special case of the general result that the peaceful outcome (N, N) is a Nash equilibrium if and only if:

$$(\pi - \beta + \sigma_1)\int_{k\in M} \alpha_1(k)\Delta(k)dk \leq c_1(C, N)$$

$$(\beta - \pi + \sigma_2)\int_{k\in M} \alpha_2(k)\Delta(k)dk \leq c_2(N, C)$$

A first step is to assume that each state cares about a mass R of issues, denoted by a compact set of points on the real line. State i cares about all issues between point a_i and point $b_i > a_i$, with $b_i - a_i = R$, but does not care about issues outside that range. To simplify notation, we also assume that all relevant issues receive equal weight $\alpha > 0$: $\alpha_i(k) = \alpha > 0$ if and only if $a_i \leq k \leq b_i$, while $\alpha_i(k) = 0$ otherwise. Hence, we can characterize the set of issues that state i cares about by a single real number v_i (to fix ideas, the mid-point in state i's set of relevant issues), which we can interpret as that state's type (or fundamental characteristics):

$$v_i = a_i + \frac{R}{2} = b_i - \frac{R}{2} \tag{5}$$

Therefore, a state of type v_i has the following preferences:[15]

$$U_i = -\int_{v_i - \frac{R}{2}}^{v_i + \frac{R}{2}} \alpha |x(k) - x_i^*(k)| dk - c_i \tag{6}$$

Let $V(i,j) \equiv |v_i - v_j|$ denote the distance between state i and state j in their fundamental characteristics.[16] We are now ready to consider conflicts between states over common issues.

2.3.1 Basic Setting

In what follows we derive the probability of conflict under two simplifying assumptions (we relax both assumptions in Appendix 1):

Assumption 1: *The extent of disagreement over all issues is constant and normalized to one - that is,* $\Delta(k) = 1$ *for all* k.

Assumption 2: *When* (N, N) *is a Nash equilibrium, the two states will always coordinate on the*

---

[15] In this simplified analysis we assume that each state is a unified agent, formed by one population with homogeneous characteristics v_j. In principle, two states can be of the same type - i.e., they may care about the identical set of issues. Here we abstract from the possibility that states may include mixed populations with different preferences over issues. However, we allow for population heterogeneity within states when computing genetic distance for our empirical analysis.

[16] Our theoretical framework abstracts from explicit geographical considerations: we study the effects of relatedness on conflict taking geographical factors as given, i.e. when considering the comparative statics of genetic distance on conflict, we are implicitly looking at states that are at a constant geographic distance from each other. However, empirically, geography and genetic distance are connected, and both have effects on the probability of conflict. We explicitly address these points in the empirical section by controlling for a wide range of geographical distance measures.

peaceful equilibrium (no coordination failure).[17]

An economic interpretation of Assumption 1 is in terms of conflict over *rival* issues. A good is rival when (a) any increase in a state's use of that good reduces the extent of the other state's use, and (b) each state's preferred outcome is to have full and exclusive use of the good. In such a context, outcome $x(k)$ can be conveniently defined as the extent to which state 1 can use the good once the dispute has been resolved, so that $x_1^*(k) = 1$ denotes state 1's ideal outcome (state 1 has full and exclusive use), and $x_2^*(k) = 0$ denotes state 2's ideal outcome (state 2 has full and exclusive use). Henceforth, $\Delta(k) = 1$ for all rival issues. Rival issues may arise when both states value the same rival good (say, a religious/cultural center or an offshore natural resource) because they have similar tastes over consumption and/or share similar production technologies. Rivalry may also emerge when the two states interact extensively with each other over an international policy issue (e.g., labor flows), and each state wants to impose its exclusive control over that issue. States may also interact with each other over non-rival issues. For example, both states may care about a common set of international public goods - e.g., security against terrorist threats, regulation of pollution or other externalities, where the use by one state would not reduce the other's ability to use the same public good, - but they may disagree about the ideal features of the public good, captured by $x(k)$ in our simplified setting. This more general case - when $\Delta(k)$ may differ from 1 - is analyzed in Appendix 1. For now we focus on the simpler case $\Delta(k) = 1$.

How does the probability of violent conflict depend on the distance in fundamental characteristics $V(i, j)$? First of all, conflict will never occur if $V(i, j) > R$. This captures the obvious but important point that two states which are very distant in the set of issues they care about will have no reason to fight. In contrast, if $V(i, j) < R$ they will share a range of common issues, and the following holds:

Remark 2: *Violent conflict (C, C) is the unique equilibrium if and only if:*[18]

$$\alpha[R - V(i, j)] > \omega \tag{7}$$

---

[17] This assumption is equivalent to limiting the analysis to Coalition-Proof Nash Equilibria, as defined in Bernheim, Peleg and Whinston (1987).

[18] For $v_i \leq v_j$, the common range includes all points between $v_j - \frac{R}{2}$ and $v_i + \frac{R}{2}$, and (C, C) is the unique equilibrium for $\int_{v_j - \frac{R}{2}}^{v_i + \frac{R}{2}} \alpha dk > \omega$, (an immediate application of Remark 1). By the same token, if $v_i \geq v_j$, (C, C) is the unique equilibrium for $\int_{v_i - \frac{R}{2}}^{v_j + \frac{R}{2}} \alpha dk > \omega$.

This simple inequality illustrates a key result: for a given range R of common issues, populations which are more distant in preferences over relevant issues are less likely to go to war with each other. In particular, if ω is a random variable distributed uniformly between 0 and $\overline{\omega}$, we have:[19]

Proposition 1: *The probability of conflict between state i and state j is:*

$$\text{Prob}(Conflict) = \frac{\alpha}{\overline{\omega}}[R - V(i,j)] \tag{8}$$

That is, conflict is increasing in the extent states care about specific common issues (α) and the range of common issues each state cares about (R), decreasing in the relative costs to start a conflict ($\overline{\omega}$), and decreasing in the distance between the two states' fundamental characteristics $V(i,j)$.

2.3.2 The Role of Relatedness

We now go a step further and derive the relationship between probability of conflict and explicit measures of long-term relatedness (genetic distance). If preferences are transmitted intergenerationally across populations (biologically and/or culturally) with variation, populations that are more closely related will be more likely to care about the same issues. This can be illustrated with a simple model of the vertical transmission of characteristics. Assume that in period t a population i inherits its type v_{it} from an ancestor population with type v_{it-1}, with variation captured by a random shock ε_{it}:

$$v_{it} = v_{it-1} + \varepsilon_{it} \tag{9}$$

Without loss of generality, consider only two periods, and assume that ε_{it} follow a simple random walk, taking value $\varepsilon > 0$ with probability 1/2 and $-\varepsilon$ with probability 1/2 (with shocks independently distributed across different populations). Let $g(i,j)$ ("genetic distance") denote the number of periods since two populations have shared common ancestors (in the empirical analysis, we use F_{ST} genetic distance, a measure that is linear in the time since two populations shared their last common ancestors). Populations at $g(i,j) = 1$ will be at $V(i,j) = 0$ with probability 1/2 and $V(i,j) = 2\varepsilon$ with probability 1/2, and hence at expected distance $E\{V(i,j) \mid g(i,j) = 1\} = \varepsilon$. By contrast, two populations at $g(i,j) = 2$ (that is, sharing a two-period-old last-common-ancestor population) will be at a higher expected distance $E\{V(i,j) \mid g(i,j) = 2\} = \frac{1}{2}2\varepsilon + \frac{1}{8}4\varepsilon = \frac{3}{2}\varepsilon$. Hence:

Remark 3: *Expected distance in inherited characteristics $V(i,j)$ is increasing in genetic distance*

---

[19] Without loss of generality, we assume values of the parameters such that $\frac{\alpha}{\overline{\omega}}[R - V(i,j)] \leq 1$.

$g(i,j)$:

$$E\{V(i,j) \mid g(i,j)=2\} - E\{V(i,j) \mid g(i,j)=1\} = \frac{\varepsilon}{2} > 0 \tag{10}$$

An immediate implication of Proposition 1 and Remark 3 is that genetic distance $g(i,j)$ is inversely related to expected probability of conflict:[20]

Corollary 1: *The expected probability of conflict between state i and state j is decreasing in genetic distance $g(i,j)$*:

$$E[\text{Prob}(Conflict) \mid g(i,j)=2] - E[\text{Prob}(Conflict) \mid g(i,j)=1] = -\frac{\varepsilon\alpha}{2\overline{\omega}} < 0 \tag{11}$$

Corollary 1 is our central theoretical result, which we test directly in the empirical section. This result highlights the logic of the link among common issues, interactions across states, long-term relatedness, and the probability of conflict. In principle, this common-issue effect could be offset by countervailing forces, such as the extent-of-disagreement effect or the coordination-failure effect, which we model and discuss in Appendix 1. However, our empirical findings document a strong and robust negative effect of genetic distance on the probability of conflict: more closely related states fight more with each other. This surprising empirical result is implied by our model. Empirically, any potential countervailing effects do not appear to be strong enough to offset the common-issue effect emerging from the simplest version of our model. Therefore, the basic theory of this section provides a consistent economic interpretation for the empirical findings.[21]

---

[20] The equation in Corollary 1 is derived under the assumption that the parameters are such that $V(i,j) < R$ for all possible realizations of the shocks. The effect of genetic distance $g(i,j)$ on the probability of conflict would remain negative if $V(i,j) > R$ - and hence no conflict were to occur - for some realizations of the shocks.

[21] The framework presented in this section could be viewed as the reduced form of a more detailed and micro-founded setting in which common issues emerge from more complex dynamic processes and decisions. For example, societies with more similar long-term characteristics might endogenously end up with more similar production systems and/or tastes and consumption patterns, which may induce them to compete over a similar set of resources. Genealogically closer populations may also face lower fixed costs to interacting with each other, and therefore have more opportunities to interact over all sorts of disputed common issues.

3 Data and Methodology

Our model shows that the degree of relatedness between populations has a positive effect on their conflict propensities due to a larger set of common issues (corollary 1). Genealogical relatedness may also affect differences in ideal points (corollary 2 in Appendix 1) and may affect the likelihood of reaching peaceful conflict resolution by facilitating coordination (corollary 3 in Appendix 1). Thus, the net effect of relatedness on conflict is a priori ambiguous. 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 a wide range of 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).[22] We start from a discrete indicator of the intensity of a bilateral conflict between countries i and j in year t. 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 taking a value of 1 if the intensity of militarized conflict is equal to or greater than 3. Our main dependent variable is this binary indicator of conflict, denoted C_{ijt}. We separately examine the determinants of the intensity of conflict, as well as the determinants of war (corresponding to a conflict intensity of 5). The database includes several other useful bilateral variables such as war casualties, 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. We make use of these variables in the analysis below.

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 only a short overview. 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 popula-

---

[22] See also Jones et. al. (1996) and Faten et al. (2004).

tions 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.[23]

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 $n = 1...N$. 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:

$$FST_{ij}^{W} = \sum_{m=1}^{M} \sum_{n=1}^{N} (s_{im} \times s_{jn} \times d_{mn}) \tag{12}$$

where s_{km} is the share of group m in country k, d_{mn} is the F_{ST} genetic distance between groups m and n. This 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 91.9%), so for practical purposes it does not make

---

[23] 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.

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.[24]

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. Hence, for instance, in the 1500 match, Australia is matched to Aborigines. Genetic distance between countries using the 1500 match can be used as an instrument for current genetic distance.[25]

3.3 Summary Statistics

Table 1 and 2 provide summary statistics that give a sense of the data and provide clues concerning the relationship between conflict and relatedness.[26] The baseline sample is an unbalanced panel of $517,251$ observations covering $13,575$ country pairs, based on 176 underlying countries, with yearly observations from 1816 to 2000. Table 1 displays the means of genetic distance, geodesic distance and a dummy variable for contiguity between the two countries in a pair, conditional on the intensity of conflict. The mean of genetic distance when there is no militarized conflict (0.102) is greater than at any other level of the conflict intensity indicator (for hostility levels ranging from 2 to 5, the mean of genetic distance ranges from 0.050 to 0.063), consistent with Corollary 1. Somewhat surprisingly, a relatively small portion of full-fledged wars (18%) occur between contiguous countries, and the mean geodesic distance separating countries at war is relatively high

---

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

[25] Since we do not have detailed data on ethnic composition going back to 1500, the corresponding match only refers to plurality groups. The matching of 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, likely reducing 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 0.714 in our baseline sample.

[26] Appendix 2 provides further summary statistics for the main variables in our study, in the form of means and correlations, to aid in the interpretation of our empirical results.

(5, 562 km). Overall, only 34% of conflicts (hostility levels equal to 3, 4 and 5) occur between contiguous countries.

Table 2 shows the conditional frequency of both wars and conflicts. Wars are rare occurrences, as only 1, 010 pair-year observations are characterized as wars, out of more than half a million observations. Over a quarter of these wars occurred between countries in the bottom decile of genetic distance, and almost half of all wars occurred in pairs in the bottom quartile. Only 44 wars were observed in pairs in the top quartile, of which 32 involved South Africa as one of the combatants. While South Africa is characterized as genetically distant from European populations due to the large African majority, a historical examination of wars involving South Africa reveals that the wars were spurred mainly by conflicts over issues separating European powers and South Africa's European power elite. In sum, there are very few wars between genetically distant populations in our sample. Even wars occurring across large geographic distances typically involve mostly genetically similar participants - for instance it is still the case that almost 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 of these conflicts (3, 728 versus 1, 010), 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.

3.4 Empirical Specification

While these summary statistics are an informative starting point, we turn to a more formal regression setup, allowing us to control for a wide range of 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. First, we collapse the panel into a single cross-section. 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 2000 time period. The baseline cross-sectional regression equation is:

$$C_{ij} = \beta_1 X_{ij} + \beta_2 FST_{ij}^W + \eta_{ij} \tag{13}$$

where C_{ij} is an indicator taking a value of 1 if a pair was ever in a conflict or war over the sample period, 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) are central control variables in our analysis. The baseline panel regression is:

$$C_{ijt} = \gamma_1 X_{ijt} + \gamma_2 FST_{ij}^W + \varepsilon_{ijt} \tag{14}$$

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 0.404).

Equations (13) and (14) 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 only 0.721%, to improve the readability of the marginal effects we multiplied all of them by 100 in all tables.

4 Empirical Results

4.1 Cross-Sectional Estimates

Table 3 presents baseline estimates of the coefficients in equation (13). 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, with a one standard deviation change in genetic distance (0.067) associated with a 68.81% decline in the percentage probability of ever having experienced a conflict (in the cross section, 5.65% of pairs were ever in a conflict between 1816 and 2000). 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.[27] These measures usually bear the expected signs (more distance, less conflict), and their inclusion reduces the effect of genetic distance.[28] However, this effect remains negative and highly significant statistically. Its magnitude is still 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 potential endogeneity of genetic distance. One 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. Another issues is measurement error stemming from imperfect matches of genetic groups to current populations and countries, leading to probable attenuation bias. To address these issues, 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 2000. Moreover, matching countries to genetic groups is

---

[27] 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).

[28] Similarly, excluding genetic distance from the baseline specification generally raises the magnitude of the geographic effects, particularly that of log geodesic distance (results are available upon request). Thus, the exclusion of relatedness from past empirical specifications seeking to explain conflict led to overstating the quantitative impact of geographic factors.

much more straightforward for 1500 for two reasons. First, Cavalli-Sforza et al. (1994) explicitly collected data for populations as they were in 1492. 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 both with and without instrumenting, keeping in mind that non-instrumented probit estimates of the effect of genetic distance are likely to be an understatement of the true magnitude.

The remaining columns of Table 3 conduct a variety of additional tests. In column 4 we isolate countries that are non-contiguous. In the baseline sample, 34% of conflicts occur between contiguous countries, and isolating pairs composed of non-contiguous countries is a further way to ensure that genetic distance does not simply capture the effect of geographic proximity. 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 conflict. This reinforces our confidence that the effect is not driven by geographic distance or other possibly omitted factors specific to contiguous countries.

Finally, we consider the determinants of wars rather than conflicts more broadly (column 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 2000. 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.

4.2 Baseline Panel Estimates

For the rest of this paper, we focus on panel estimates. This is justified by the need to control for time varying factors, such as the extent to which the countries in a pair are democratic, the intensity of trade links between pairs, income differences, and variables reflecting other potential determinants of conflicts (the presence of alliances, the number of other conflicts occurring at the time, etc.). Table 4 presents baseline estimates from the panel specification of equation (14) over the period 1816-2000 (later we will turn to estimates for the post-1950 period, for which more time-varying controls are available).

The results are consistent with those already discussed: genetic distance significantly reduces the probability of conflict, although the magnitude is smaller in the panel than in the cross-section. In the specification with the broadest set of controls (column 3), the standardized magnitude of the effect is 8.52% of the mean percentage probability of conflict (in the panel, this mean probability is 0.721%). Another noteworthy finding in this table relates to the central tenet of liberal peace theory, namely the idea that democracies tend not to go to war with each other. A dummy variable equal to 1 if both countries are democracies (defined as a combined Polity score greater than 5) has a negative and highly significant marginal effect, with roughly the same magnitude as that of genetic distance. Column (3) includes other time-varying controls such as the number of peaceful years in the pair prior to the current year, the number of wars taking place globally at time t, and whether the two countries are members of the same alliance. All of these bear coefficients with the expected signs. We continue to condition on the full set of controls in the baseline specification of column 3 in all the regressions that follow.

To assess the sensitivity of our results to assumptions about the distribution of the error term, we considered a logit rather than a probit estimator. The findings do not change (column 4). We continue to use a probit estimator in the rest of this paper, because it is computationally faster. Finally, in column 5, we instrument for current genetic distance using genetic distance between populations as they were in 1500. As in the cross-section, the effect of genetic distance rises by over 50% - to 13.19% - relative to the estimate of column (3).

4.3 Estimates by Type of Conflict

In this subsection, we examine whether the effect of genetic distance differs by type of conflict, focusing on territorial versus non-territorial conflicts. We exploit information available in the

Correlates of War database on the type of dispute. The database codes, for each country in a conflict pair, whether it seeks a revision in the territorial *status quo*.[29] We define a territorial conflict as one for which either country seeks a territorial revision either as the most inportant or the second most important rationale for the dispute. This coding rule is the most conservative in characterizing a conflict as non-territorial.[30]

Table 5 presents the results, using both probit and IV probit estimators on the panel dataset, using the same baseline specification with the broadest set of controls. The first two columns repeat the baseline regressions on all types of conflicts, to facilitate comparisons. Columns (3) and (4) focus on the determinants of territorial conflicts. The effect of genetic distance, while it remains negative and statistically significant at the 1% level, is much reduced in magnitude. In contrast, for non-territorial conflicts (columns 5 and 6), the standardize effect is larger than in the baseline. Thus, interestingly, the effect of genetic distance on the probability of conflict seems to arise, in quantitative terms, mostly from non-territorial conflicts.

These results provide further insights on the potential mechanisms through which relatedness may affect conflict. Most importantly, we can rule out that the effect of genetic distance on conflict is working *exclusively* through geographic/territorial channels. The fact that the effect is stronger for non-territorial conflict is consistent with our broad theoretical framework, where the effect stems from disputes over sets of common issues, which may or may not be related to territorial expansion. In addition, these results confirm that the effect of relatedness on conflict is unlikely to capture geographic factors remaining after inclusion of numerous geographic controls. One would expect that geographic factors should play a larger role as a determinant of territorial conflicts, so a high and significant impact of genetic distance on non-territorial conflicts is further evidence that relatedness exerts an effect separate from that of geographic distance. Finally, to the extent that conflict is motivated by goals of territorial conquest, pairs with a history of persistent territorial conflicts may have experienced some degree of population mixing, raising the possibility of genetic distance being endogenous to territorial conflict. Such endogeneity concerns are alleviated by the fact that genetic distance mostly affects non-territorial conflicts.

---

[29] Non-territorial issues include a desire to change the other country's regime or to change the other country's policies. See also Vasquez and Henehan (2001) for more details on the Correlate of Wars coding of revisionist states and territorial conflicts. These codings have been widely used in the political science literature on interstate conflicts.

[30] With this definition, our original conflict variable is the sum of the indicators for territorial and non-territorial conflicts. Of the 3,728 conflicts in our baseline sample, 1,096 were coded as territorial and 2,632 as non-territorial.

4.4 Estimates Across Time and Space

To examine if specific historical periods account for the finding of a negative effect of relatedness on conflict, we broke down the sample by time period. Results are presented in Table 6. We find that results are remarkably robust across time periods. The coefficient on genetic distance is negative and roughly of the same magnitude whether considering the pre- or post-1900 periods. The coefficient for the pre-1900 period is not statistically significant, perhaps because there are many fewer observations in the early periods (only 799 country pairs as opposed to $13,175$ for the broader sample), and few observations with conflict (436 out of a total of $3,728$ conflicts in the broader sample). Focusing on the 20th century, the effect is particularly pronounced and significant for the post 1946 period - 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.[31] In fact, our finding holds even after the end of the Cold War (column 6). The coefficient is negative whatever the subperiod under consideration.

In further regressions that are available upon request, we broke down the sample across space, by continent. We again uncovered a negative effect of genetic distance on conflict within every continent. The results were particularly striking for Europe, for which a separate matrix of F_{ST} genetic distances is available.[32] Despite the paucity of observations (only 291 country pairs), the effect of genetic distance remained negative and significant at the 5% level, and slightly larger in standardized magnitude compared to worldwide results. While the relationship between war and relatedness holds strongly within Europe, this continent does not drive the results obtained in the worldwide regressions, as they hold even when all European countries are excluded from the sample. The effect was consistently negative for all other continents where conflicts occurred.[33] Overall, the regional breakdown suggests that the negative effect of relatedness on war is remarkably consistent

---

[31] $2,053$ observations involve militarized conflicts in the post 1946 period, while the 1939-1946 period involved 634 bilateral conflict-years, or 17% of the total number of observations with conflicts between 1816 and 2001.

[32] 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.

[33] The coefficient was negative and significant at the 10% level for Asia and Africa, and negative but insignificant for the Americas. The number of intracontinental interstate conflicts experienced by these continents were 787 (Asia), 252 (Africa) and 433 (Americas). There were no conflicts within Oceania in our baseline sample.

across space.

4.5 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, 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).[34] 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.[35]

Prior to discussing the results, we briefly describe how these measures were constructed. To capture linguistic distance, we used 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).[36] Using data on the distribution of each linguistic group within and across

---

[34] 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.

[35] Pairwise correlations between measures of genetic, linguistic and religious distances appear in Appendix 2, panel b. These correlations are generally positive, as expected, but not very large. For instance, the correlation between F_{ST} genetic distance and weighted linguistic distance is 0.164. Religious distance bears a correlation of 0.544 with linguistic distance, and 0.210 with genetic distance.

[36] 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

countries, from the same source, we again computed a measure of the number of common nodes shared by languages spoken by plurality groups within each country in a pair. We also computed 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 12).[37] Following Fearon (2003), we transformed these measures so that they reflect linguistic distance (LD) rather than similarity, and are bounded by 0 and 1:

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

To measure religious distance we followed 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 matched religions to countries using Mecham, Fearon and Laitin's (2006) data on the prevalence of religions by country and transformed the data in a manner similar to that in equation (15), again computing plurality and weighted distances separately.

Table 7 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 $77,081$ observations, or almost 15% 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 related with conflict, though the effect is only significant at the 7.6% level, and its

---

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.

[37] 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.

significance level drops to 13% when including linguistic distance along with religious distance.[38] This latter finding, while weak, 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.[39]

4.6 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 8). We then seek to evaluate the effect of genetic distance on the intensity of conflict, rather than on a binary indicator of conflict incidence (Table 9).

We first isolate countries that are non contiguous. In the baseline sample, 34% of conflicts occur between contiguous countries, and isolating pairs composed of non-contiguous countries is a further way to control for geographic proximity. The standardized effect of genetic distance actually *rises* modestly, as a one standard deviation increase in genetic distance is associated with a 9.41% decrease in the mean probability of conflict (versus 8.52% in the baseline regression). This reinforces our confidence that the effect is not driven by geographic distance or other possibly omitted factors specific to contiguous countries.

In columns (3) through (5) of Table 8 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 are contiguous, for pairs that are geographically proximate (i.e. countries are either contiguous or separated by a distance less than $2,500$ km), or for pairs that include a major power. 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 0.095, 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. In

---

[38] This result contrasts with that in Henderson (1997), who found evidence that religious *similarity* was negatively related to conflict. The difference may stem from a much bigger sample in our work, as well as our inclusion of a much broader set of controls (Henderson only controlled for contiguity).

[39] The estimated effects of religious and linguistic distance do not change much when genetic distance is excluded from the regression, although religious distance becomes more significant statistically. Corresponding results are available upon request.

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.[40]

Table 9 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. 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. This result rationalizes our focus on a bilateral measure of conflict rather than on the continuous measure. In line with results in Tables 3 and 4, instrumenting for genetic distance based on the current match of populations to countries using genetic distance based on the 1500 match increases the estimated magnitude of the effect by 64% (column 3).

In columns (4) and (5) we consider the determinants of war casualties. We find that genetic distance reduces 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. Our last test (column 6) is to redefine the dependent variable as a binary indicator of war, i.e. a dummy variable taking on a value of one if conflict intensity is 5 (corresponding to conflicts with more than $1,000$ total battle deaths). Genetic distance reduces the propensity for war in a statistically significant way: a standard deviation increase in genetic distance reduces the probability of full-blown war by 2.96% of this variable's mean, an effect quantitatively smaller than that on conflict more broadly (the underlying probability of a country pair-year being at war in our baseline sample is relatively low, on the order of 0.195%). These results are consistent with those found with the cross-sectional approach of Table 3.

To summarize, the effect of genetic distance is very robust to using alternative measures of conflict. We also uncover little evidence that genetic distance affects the intensity of conflict conditional on a conflict occurring.

---

[40] In further tests available upon request, we allowed for nonlinear effects of geographic distance to capture the 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 found to be insignificant.

4.7 Analysis for the 1950-2000 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 1950-2000 period for which various measures of trade and income are available.

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 a recent paper, 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, to the extent that genetic distance and trade are correlated.

We obtained the same data on bilateral and multilateral trade openness used in MMT's paper, and included their measures of trade in our baseline specification.[41] 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 1950-2000. We recover the pattern of coefficients on the trade terms as the one reported 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. Our finding lend further support to liberal peace theory, as recently amended by MMT. 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 6.612% of this variable's mean. Adding the trade terms in column (2), this effect falls further, but remains negative and highly

---

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

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 the 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 log income differences (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).[42] The coefficient on income differences is positive and significant, indicating 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. 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.

4.8 Analysis of UN voting patterns

In our theoretical framework, Corollary 2 suggests that one way relatedness could affect conflict is through its effect on the degree of similarity in countries' ideal points over non rival issues. Stated simply, related populations may have more or less similar preferences over sets of international issues, quite apart from the effect of relatedness on the range of issues relevant to the pair, stemming from past interactions. In the theory section, we remained agnostic as to the possible direction of this relationship. In this subsection we seek to uncover empirically the direction of the effect by analyzing the degree of countries' similarity in stated preferences over global diplomatic issues. To do so, we use data on their voting patterns at the UN General Assembly. The data comes from Gartzke (2006), who states that "dozens or hundreds of resolutions appear in each session of the

---

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

General Assembly." Most of these votes constitute symbolic position taking by UN members, who usually do not have a direct stake in the issue they vote on. Another advantage of this data is that all UN members take positions (including abstaining) on a constant set of issues.

Based on data on votes themselves, Gartzke constructed an index of the "affinity of nations", which is simply the bilateral correlation of votes for each country pair in a given year. The measure ranges from -1 to 1 and is available from 1946 to 2002. Two separate indices are available depending on whether abstentions are considered a form of position taking, or excluded. We use both indices as dependent variables to examine the effect of genetic distance on the degree of similarity in preferences over diplomatic issues considered at the UN General Assembly. We maintain the same baseline specification used to estimate the determinants of conflicts (Table 3, column 3), regressing UN vote correlation indices on genetic distance, geographic distance and other controls.

Estimates suggest that genetic distance is positively associated with UN vote correlations. That is, countries that are more related have more *different* preferences over issues arising at the UN Assembly. Column (2) of Table 11 shows this is the case unconditionally. The effect remains positive and significant when including a set of geographic and historical controls (columns 2 and 3). The effect remains when considering only the 1990-2000 time period where votes were less likely to be aligned with the major geopolitical blocs of the Cold War era. The effect is also robust to excluding abstensions from the calculation of UN vote correlations. In terms of magnitude, using the baseline regression of column (3), a one standard deviation increase in genetic distance is associated with an increase in the UN vote correlation equal to 10.10% of this variables standard deviation, i.e. the standardized beta is 10.10%. This standardized measure of magnitude rises to 11.49% when excluding abstentions.

To summarize, this evidence suggests that any positive effect of relatedness on conflict arising from the role of past interactions in generating grievances is likely to be reinforced by the negative effect of relatedness on preference similarity (in the parlance of our model, δ is negative).

5 Conclusion

In this paper, we examined the empirical relationship between the occurrence of international conflicts and the degree of relatedness between countries. We found that populations that are genetically closer are more prone to engage in militarized conflicts with each other, even after controlling for a wide set of measures of geographic distance, income differences, and other factors affecting conflict, including measures of bilateral and multilateral trade and differences in democracy levels. The effect of genetic distance on conflict is large in magnitude, and even higher when instrumenting using historical genetic distance. We also provided a theoretical model of conflict and relatedness that is consistent with these results. In the simplest version of our model, populations that share a more recent common history have had less time to diverge in preferences and characteristics that determine the set of common issues they care about, and over which they are prone to fight.

To our knowledge, this is the first paper that documents a link between genetic distance and international conflict, and provides an interpretation in terms of historical relatedness. As we have discussed in the introduction, our results provide strong evidence against the primordialist view that ethnic dissimilarity should breed war and plunder.

More broadly, this paper is part of a growing literature in political economy focusing on the effects of long-term historical variables on political, economic and institutional outcomes, both theoretically and empirically. It would be interesting to link our approach to the extensive literature on ethnic fractionalization and polarization within countries (see Alesina et al. 2003, Fearon, 2003) and to study the effects of long-term genealogical relatedness across groups on civil conflicts and other intrastate outcomes. A positive relationship between relatedness and conflict within states would be consistent, for example, with the finding in Fearon and Laitin (2003) that ethnic fractionalization and civil wars are unrelated. Further research on this question should focus on reliable subnational data on inter-group relatedness.[43]

---

[43] Another area of research where our approach could be fruitful is the study of national formation and breakup, and their connections with international conflict (Spolaore, 2004; Alesina and Spolaore, 2005, 2006) and civil conflict (Spolaore, 2008). A first attempt to use genetic distance to study the political economy of national borders is Desmet et al. (2007).

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 (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.

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

Bernheim, B., B. Peleg, and M. Whinston (1987) "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory*, 42(1): 1-12.

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.

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. Ortuno-Ortin and S. Weber (2006), "Nation Formation and Genetic Diversity," CEPR Discussion Paper 5918.

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.

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 (2006), "Economics of Conflict: An Overview," in K. Hartley and T. Sandler (eds) *Handbook of Defense Economics*, Volume 2, North Holland, Amsterdam.

Gartzke, E. (2006), "The Affinity of Nations Index, 1946-2002," Version 4.0, University of California San Diego, http://dss.ucsd.edu/~egartzke/ (accessed 06/06/2009).

Gartzke, E. and K. S. Gleditsch (2006), “Identity and Conflict: Ties that Bind and Differences that Divide,” *European Journal of International Relations*, 12(1): 53-87.

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

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

Heston, A., R. Summers, and B. Aten, "Penn World Table Version 6.1," *Center for International Comparisons at the University of Pennsylvania* (CICUP), 2002.

Hirshleifer, J. (1991), "The Technology of Conflict as an Economic Activity," *American Economic Review,* 81(2): 130-134.

Hirshleifer, J.(1995), "Anarchy and its Breakdown," *Journal of Political Economy,* 103(1): 26-52.

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.

Keegan, J. (1984), *A History of Warfare*, New York: Vintage.

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.

Merton, R. (1957), *Social Theory and Social Structure*, Glencoe: Free Press.

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

Nye, J. S. (2008), *Understanding International Conflicts. An Introduction to Theory and History,* Longman, 7th edition.

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.

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.

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.

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

Spolaore, E. (2004), "Economic Integration, International Conflict and Political Unions," *Rivista di Politica Economica,* 3-50.

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

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

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.

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.

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.

Appendix 1: Disagreement and Coordination Failure

A1. Extent of Disagreement

We now relax Assumptions 1 and extend the analysis to the more general case in which the extent of disagreement $\Delta_{ij}(k) = |x_i^*(k) - x_j^*(k)|$ is not necessarily equal to 1 for all issues, but may vary in functions of the inherited characteristics of the two states.[44] A priori, the relationship between inherited characteristics and extent of disagreement can go either way. More closely related populations may face less disagreement over non-rival common issues (e.g., specific international public goods), which would reduce the probability of conflict, *ceteris paribus*. On the other hand, closely related populations that care a lot about the same non-rival issue may also have strongly divergent preferences over the details of how the issue should be settled, and hence be farther away in their ideal points (for example, related populations that care about the same religious/cultural issue may also greatly diverge in their ideal outcomes). Here we present a direct formalization linking the extent of disagreement to the distance in inherited characteristics, and study the relation between probability of conflict and genetic distance in this more general setting (in Section 4 we presented some evidence on voting patterns at the United Nations that empirically showed an inverse relation between relatedness and the extent of agreement over international issues).

Assume that for any set of issues between any two points on the real line, a fraction ρ is rival and a fraction $(1-\rho)$ is non-rival, and that the extent of disagreement over non-rival common issues between state i and state j may depend on the distance in inherited characteristics $V(i,j)$:

$$\Delta_{ij}(k) = \Delta_0 + \delta V(i,j) \tag{16}$$

where $\Delta_0 \geq 0$ and δ is a parameter measuring the relation between distance $V(i,j)$ and disagreement $\Delta_{ij}(k)$, and k is a non-rival issue. The other assumptions of our model are maintained. All states at a distance $V(i,j) > R$ share no common issues, and hence face no conflict. For states at a distance $V(i,j) \leq R$, violent conflict is the only equilibrium if and only if $\alpha[\rho + (1-\rho)\Delta_0][R - V(i,j)] > \omega$. To simplify notation, we assume that all relevant issues share the same α (the results would not change qualitatively if we assume that relevant rival issues enter the utility function with parameter

---

[44] Clearly if $\Delta_{ij}(k)$ is independent of $V(i,j)$ for all issues, the qualitative results from the basic setting will not be affected.

$\alpha_r > 0$ while non-rival issues enter with parameter $\alpha_{nr} > 0$). Assuming again that ω is a random variable distributed uniformly between 0 and $\overline{\omega}$, we now have:

Proposition 2: *The probability of conflict between the two states is given by:*

$$\text{Prob}(Conflict) = \frac{\alpha}{\overline{\omega}} \left\{ [\rho + (1-\rho)\Delta_0]R - [\rho + (1-\rho)(\Delta_0 - \delta)]V(i,j) - (1-\rho)\delta[V(i,j)]^2 \right\} \quad (17)$$

A negative relationship between probability of conflict and distance $V(i,j)$ *holds (i.e.,* $\frac{d\,\text{Prob}(Conflict)}{dV(i,j)} < 0$*) if* δ *is small enough:*

$$\delta < \frac{\rho + (1-\rho)\Delta_0}{(1-\rho)[1 - 2V(i,j)]} \quad (18)$$

The above inequality is always satisfied if $\delta \leq 0$. If $\delta > 0$, the inequality is more easily satisfied the larger the fraction of rival issues ρ, and the larger the extent of disagreement which is independent of distance Δ_0. An analogous condition can be stated in terms of expected probability of conflict and genetic distance. By taking expectations of the above Prob($Conflict$), using the facts that $E\{V(i,j) \mid g(i,j) = 1\} = \varepsilon$, $E\{V(i,j)^2 \mid g(i,j) = 1\} = 2\varepsilon^2$, $E\{V(i,j) \mid g(i,j) = 2\} = \frac{3}{2}\varepsilon$ and $E\{V(i,j)^2 \mid g(i,j) = 2\} = 4\varepsilon^2$, we have:

$$\begin{aligned} E[\text{Prob}(Conflict) \mid g(i,j) &= 2] - E[\text{Prob}(Conflict) \mid g(i,j) = 1] \quad (19) \\ &= -\frac{\alpha}{\overline{\omega}}\{[\rho + (1-\rho)(\Delta_0 - \delta)]\varepsilon + 2(1-\rho)\delta\varepsilon^2\} \end{aligned}$$

which immediately implies:

Corollary 2: *Expected conflict is decreasing in genetic distance (i.e.,* $E[\text{Prob}(Conflict) \mid g(i,j) = 2] < E[\text{Prob}(Conflict) \mid g(i,j) = 1]$*) if:*

$$\delta < \frac{\rho + (1-\rho)\Delta_0}{(1-\rho)\varepsilon(1-2\varepsilon)} \quad (20)$$

Consequently, an inverse relationship between conflict and genetic distance is consistent with a small or negative effect of distance $V(i,j)$ on the extent of disagreement over non-rival issues, and/or with a predominance of rival issues in international disputes.

A2. Coordination Failure

Our framework can be further extended by relaxing Assumption 2 - that is, by allowing for the possibility of coordination failure. As we have noted in Section 2, (C,C) is always a Nash equilibrium . Nonetheless, both states would be better off with a peaceful negotiation than with violent conflict (C,C), because of the costs of war, which are not borne in a peaceful outcome. As we have seen, if (N,N) is a Nash equilibrium, it is the unique coalition-proof Nash equilibrium, as defined by Bernheim, Peleg and Whinston (1987). But what if states fail to coordinate on such superior (N,N) equilibrium, and end up in the inferior (C,C) equilibrium, even when the conditions for a peaceful equilibrium are satisfied? In principle, such coordination failure could be more likely across populations that are genealogically more distant, since their norms, habits, languages etc. would tend to be more different, and they might therefore find communication and coordination more difficult. Then, such coordination failure effect would reduce the negative correlation between genetic distance and probability of conflict (in contrast, if coordination failure were more likely between more closely related populations, the effect of relatedness on conflict would be strengthened). A formalization of these ideas is provided as follows. Let $\chi(i,j)$ denote the probability that state i and state j would fail to coordinate on the peaceful outcome when it is an equilibrium, and assume that such coordination failure is a function of the distance in fundamental characteristics, measured by $V(i,j) = v_i$ and v_j:

$$\chi(i,j) = \chi_0 + \theta V(i,j) \tag{A1}$$

with $\chi_0 \geq 0$ and θ is a parameter measuring the relation between distance $V(i,j)$ and probability of coordination failure $\chi(i,j)$. Also assume that all parameters satisfy the appropriate restrictions to ensure that $0 \leq \chi(i,j) \leq 1$. Therefore, for $V(i,j) \leq R$, and assuming again that ω is a random variable distributed uniformly between 0 and $\overline{\omega}$, we have:

Proposition 3: *The probability of conflict between the two states when all common issues are rival ($\rho = 1$) is:*

$$\begin{aligned} \text{Prob}\{Conflict\} &= \chi(i,j)\Pr ob[\alpha[R - V(i,j)] \leq \omega] + \text{Prob}[\alpha[R - V(i,j)] > \omega] \qquad \text{(A2)} \\ &= \chi(i,j) + [1 - \chi(i,j)]\frac{\alpha}{\overline{\omega}}[R - V(i,j)] \end{aligned}$$

The probability of conflict is decreasing in distance $V(i,j)$ if:

$$\theta < \frac{1 - \chi_0}{\dfrac{\overline{\omega}}{\alpha} - R - 2V(i,j)} \tag{A3}$$

An analogous condition holds for the more general case $\rho \leq 1$ (see generalization below).

The above inequality always holds for $\theta < 0$. For a positive θ, it is more easily satisfied for smaller χ_0, larger α and R, and larger $V(i,j)$.

An analogous condition holds regarding the relation between expected probability of conflict and genetic distance:

Corollary 3: *Expected conflict is decreasing in genetic distance (i.e.,* $E[\text{Prob}(Conflict) \mid g(i,j) = 2] < E[\text{Prob}(Conflict) \mid g(i,j) = 1]$*) if:*

$$\theta < \frac{1 - \chi_0}{[\frac{\overline{\omega}}{\alpha} - R - 2\varepsilon]\varepsilon} \tag{A4}$$

Consequently, if observed conflict partly stems from coordination failure, an inverse relationship between conflict and genetic distance (as the one we actually observe in the data) is consistent with a small (or even negative) effect of relatedness on the probability of coordination failure (low θ).

Generalization of Proposition 3 for $\rho \leq 1$: For the more general case $\rho \leq 1$, we have:

$$\text{Prob}\{Conflict\} = \chi(i,j) + [1 - \chi(i,j)]\,\text{Prob}\{C, C\ Unique\ Eq\} \tag{A5}$$

where:

$$\text{Prob}\{C, C\ Unique\ Eq\} = \frac{\alpha}{\overline{\omega}}\left\{[\rho + (1-\rho)\Delta_0]R - [\rho + (1-\rho)(\Delta_0 - \delta)]V(i,j) - (1-\rho)\delta[V(i,j)]^2\right\} \tag{A6}$$

By substituting $\chi(i,j) = \chi_0 + \theta V(i,j)$ and taking the derivative of the probability of conflict with respect to distance $V(i,j)$, we have the following generalization of Proposition 3:

The effect of distance $V(i,j)$ *on the probability of conflict is negative if:*

$$\theta < \frac{(1-\chi_0)\left\{\rho + (1-\rho)(\Delta_0 - \delta) + 2(1-\rho)\delta V(i,j)\right\}}{\frac{\overline{\omega}}{\alpha} - [\rho + (1-\rho)\Delta_0]R - 2[\rho + (1-\rho)(\Delta_0 - \delta)]V(i,j) - 3(1-\rho)\delta[V(i,j)]^2} \tag{A7}$$

which reduces to the condition $\theta < \frac{1 - \chi_0}{\frac{\overline{\omega}}{\alpha} - R - 2V(i,j)}$ for $\rho = 1$ (an analogous condition can be obtained for the relation between expected probability of conflict and genetic distance, along the lines illustrated above).

Appendix 2 – Summary statistics and correlations for the main variables in the analysis

Panel a – Summary Statistics

Variable	# Obs.	Mean	Std. Dev.	Min	Max
Conflict (%)	517,251	0.721	8.459	0	100
War (%)	517,251	0.195	4.415	0	100
Fst genetic distance	517,251	0.102	0.066	0	0.355
Log geodesic distance	517,251	8.690	0.816	2.349	9.899
Dummy for contiguity	517,251	0.026	0.159	0	1
Religious Distance Index, weighted	443,472	0.811	0.186	.089	1
Linguistic Distance Index, weighted	440,170	0.944	0.160	0	1

Panel b – Pairwise Correlations

	Conflict (%)	War (%)	FST genetic distance	Log geodesic distance	Contiguity	Religious distance	Linguistic distance
War (%)	0.519* (517,251)	1 (517,251)					
Fst genetic distance	-0.058* (517,251)	-0.026* (517,251)	1 (517,251)				
Log geodesic distance	-0.101* (517,251)	-0.033* (517,251)	0.404* (517,251)	1 (517,251)			
Dummy for contiguity	0.168* (517,251)	0.043* (517,251)	-0.148* (517,251)	-0.391* (517,251)	1 (517,251)		
Religious Distance Index, weighted	-0.020* (443,472)	0.001 (443,472)	0.210* (443,472)	0.243* (443,472)	-0.175* (443,472)	1 (443,472)	
Linguistic Distance Index, weighted	-0.035* (440,170)	-0.001 (440,170)	0.164* (440,170)	0.255* (440,170)	-0.193* (440,170)	0.544* (440,170)	1 (440,170)

(# of observations in parentheses)

Table 1 – Means of Genetic Distance, Geodesic Distance and Contiguity Conditional on the Overall Level of Hostility *

	Fst Genetic Distance	Geodesic Distance, km	Contiguity dummy
Overall Hostility Level = 0 (# of observations = 513,407)	0.102	7,635	0.024
Overall Hostility Level = 2 (# of observations = 116)	0.050	3,455	0.284
Overall Hostility Level = 3 (# of observations = 626)	0.054	3,495	0.372
Overall Hostility Level = 4 (# of observations = 2,092)	0.055	3,695	0.408
Overall Hostility Level = 5 (# of observations = 1,010)	0.063	5,562	0.182

517,251 pair-year observations from 13,575 country pairs.
* No observations involved an overall hostility level equal to 1 in the sample.
The overall hostility level is defined by COW as the maximum of each country's hostility level within a pair. Hostility levels are defined in COW as follows: 0=No hostility, 1=No militarized action, 2=Threat to use force, 3=Display of force, 4=Use of Force, 5=War.

Table 2 – Conditional Frequency of War (number of pair-year observations 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	277 (27.4%)	487 (48.2%)	178 (17.6%)	301 (29.8%)	44 (4.4%)	1,010 (100%)
Common sea / ocean = 0	170 (22.0%)	329 (42.7%)	129 (16.7%)	269 (34.9%)	44 (5.7%)	771 (100%)
Contiguity = 0	175 (21.2%)	368 (44.6%)	123 (14.9%)	291 (35.2%)	44 (5.3%)	826 (100%)
Distance > 1000 km	163 (19.5%)	349 (41.7%)	155 (18.5%)	289 (34.5%)	44 (5.3%)	837 (100%)
	Hostility Level > 3 (Conflict)					
None	1,076 (28.9%)	1,937 (52.0%)	940 (25.2%)	717 (19.2%)	134 (3.6%)	3,728 (100%)
Common sea / ocean = 0	582 (24.6%)	1,128 (47.7%)	569 (24.1%)	553 (23.4%)	114 (4.8%)	2,364 (100%)
Contiguity = 0	537 (21.9%)	1,202 (48.9%)	520 (21.2%)	616 (25.1%)	119 (4.8%)	2,457 (100%)
Distance > 1000 km	512 (18.2%)	1,210 (43.1%)	780 (27.8%)	684 (24.4%)	134 (4.8%)	2,808 (100%)

Based on 517,251 pair-year observations from 13,575 country pairs.
* 32 of the 44 cases in rows 3-6 involve South Africa as a combatant.

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

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

Robust t statistics in parentheses; * significant at 5%; ** significant at 1%. The standardized magnitude refers to the effect of a one-standard deviation increase in genetic distance as a percentage of the mean probability of conflict/war for the sample used in each regression.
Probit marginal effects are reported in all columns. For dummy variables, marginal effects are for discrete changes from 0 to 1. All marginal effects were multiplied by 100 for readability.

Table 4 – Baseline panel analysis, 1816-2001
(dependent variable: dichotomous indicator of conflict)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
	Univariate, probit	**Geography controls, probit**	**Baseline, probit**	**Baseline, logit**	**Baseline, IV probit**
Fst genetic distance	**-7.4543** **(12.297)****	**-1.3275** **(5.837)****	**-0.9313** **(8.922)****	**-0.7389** **(-6.224)****	**-1.4414** **(-6.511) ****
Log geodesic distance		-0.1577 (4.842)**	-0.0735 (4.487)**	-0.0435 (-2.964)**	-0.0531 (-2.787) **
Log absolute difference in longitudes		-0.0120 (0.579)	-0.0003 (0.029)	-0.0089 (-1.189)	-0.0085 (-0.812)
Log absolute difference in latitudes		-0.0607 (3.276)**	-0.0250 (2.927)**	-0.0249 (-3.909)**	-0.0284 (-3.166) **
1 for contiguity		0.8897 (7.713)**	0.4227 (7.760)**	0.1617 (3.675)**	0.4346 (3.716) **
Number of landlocked countries in the pair		-0.2088 (6.219)**	-0.1197 (7.553)**	-0.0875 (-6.392)**	-0.1219 (-7.012) **
Number of island countries in the pair		0.1712 (4.312)**	0.0551 (2.969)**	0.0468 (3.240)**	0.0616 (3.255) **
1 if pair shares at least one sea or ocean		0.0782 (1.900)	0.1029 (4.501)**	0.0657 (3.281)**	0.1049 (3.264) **
Log product of land areas in square km		0.0986 (13.263)**	0.0511 (15.762)**	0.0398 (12.889)**	0.0532 (11.687) **
1 if both countries are democracies (polity2>5)			-0.0935 (8.670)**	-0.0816 (-8.614)**	-0.1012 (-8.989) **
1 for pairs ever in colonial relationship			0.1478 (3.413)**	0.0708 (2.096)*	0.1541 (2.272) *
1 if countries were or are the same country			0.0444 (1.021)	0.0344 (1.031)	0.0526 (0.948)
Number of peaceful years			-0.0066 (13.545)**	-0.0074 (-14.131)**	-0.0069 (-11.182) **
Number of other wars in year t			0.0035 (16.748)**	0.0025 (9.447)**	0.0039 (9.666) **
Dummy for alliance active in year t			-0.0593 (4.686)**	-0.0450 (-5.063)**	-0.0537 (-4.591) **
Standardized magnitude, %	-68.203	-12.146	-8.521	-6.791	-13.190

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 columns (1)-(3). Logit marginal effects reported in column (4). 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 5 – Sample breakdown by type of conflict (1816-2001 panel)
(dependent variable: dichotomous indicator of various types of conflicts, as specified in the second row)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Baseline specification, probit	**Baseline specification, IV probit**	**Territorial conflicts, probit**	**Territorial conflicts, IV probit**	**Non territorial conflicts, probit**	**Non territorial conflicts, IV probit**
Fst genetic distance	**-0.9313**** **(-8.922)**	**-1.4414**** **(-6.511)**	**-0.0901**** **(-4.801)**	**-0.1485**** **(-2.836)**	**-0.7940**** **(-8.345)**	**-1.1945**** **(-6.718)**
Log geodesic distance	-0.0735** (-4.487)	-0.0531** (-2.786)	-0.0037 (-1.532)	-0.0017 (-0.552)	-0.0692** (-5.288)	-0.0517** (-3.365)
1 for contiguity	0.4227** (7.760)	0.4346** (3.716)	0.1236** (7.732)	0.1277* (2.407)	0.1454** (4.692)	0.1501** (2.933)
Standardized magnitude, %	-8.521	-13.190	-2.805	-4.623	-10.290	-15.480

Robust t statistics in parentheses (clustering at the country pair level); * 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 each of the various types of conflicts. 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.

All specifications were estimated with 517,251 observations from 13,175 country pairs.

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.

Table 6 – Sample breakdown by historical subperiod
(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	**-0.9313** **(8.922)****	**-0.8059** **(0.526)**	**-0.7590** **(8.668)****	**-0.3915** **(7.293)****	**-0.7591** **(7.372)****	**-0.3555** **(5.317)****
Log geodesic distance	-0.0735 (4.487)**	-0.3183 (2.887)**	-0.0608 (4.247)**	-0.0216 (2.919)**	-0.0717 (4.225)**	-0.0081 (1.449)
1 for contiguity	0.4227 (7.760)**	0.8359 (3.759)**	0.3893 (7.645)**	0.2395 (7.411)**	0.5238 (7.235)**	0.1203 (4.851)**
# observations	517,251	32,292	484,959	423,790	330,365	139,159
# of pairs	13,175	799	13,175	13,175	10,397	13,175
Standardized magnitude, %	-8.521	-2.666	-7.413	-5.454	-6.736	-7.062

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 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.

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.

Table 7 – Adding other measures of historical distance
(dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(4)
	Baseline specification (common sample)	**Add linguistic distance**	**Add religious distance**	**Add religious and linguistic distances**
Fst genetic distance	**-1.1697 (8.613)****	**-1.1717 (8.608)****	**-1.1466 (8.501)****	**-1.1461 (8.495)****
Log geodesic distance	-0.0920 (4.301)**	-0.0921 (4.284)**	-0.0936 (4.390)**	-0.0936 (4.390)**
1 for contiguity	0.5621 (7.894)**	0.5516 (7.853)**	0.5219 (7.676)**	0.5224 (7.719)**
Linguistic Distance Index, weighted	-	-0.0298 (0.539)	-	0.0028 (0.043)
Religious Distance Index, weighted	-	-	-0.0872 (1.777)	-0.0882 (1.518)
Standardized magnitude, %	-9.096	-9.112	-8.916	-8.913

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.

All specifications were estimated with 440,170 observations from 10,021 country pairs.

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, nog product of land areas in square km, dummy for both countries are democracies (polity2>5), dummy for pairs ever in colonial relationship, dummy for 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.

Table 8 - Nonlinearities and sample splits
(dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	**Excluding contiguous pairs**	**Contiguity interaction**	**Proximity interaction**	**Major power interaction**	**Spline**	**Quadratic**
Fst genetic distance	**-0.9313** **(8.922)****	**-0.6978** **(8.332)****	**-0.9706** **(8.982)****	**-0.9852** **(8.858)****	**-0.9319** **(8.882)****	**-1.0774** **(3.882)****	**-0.3287** **(0.965)**
Log geodesic distance	-0.0735 (4.487)**	-0.0865 (6.131)**	-0.0750 (4.703)**	-0.0700 (4.269)**	-0.0592 (4.171)**	-0.0731 (4.414)**	-0.0743 (4.652)**
1 for contiguity	0.4227 (7.760)**		0.3232 (4.976)**	0.4030 (7.639)**	0.4536 (8.202)**	0.4261 (7.733)**	0.4064 (7.742)**
Fst Genetic Distance * Contiguity			0.5419 (1.177)				
Fst Genetic Dist ance * Proximity				0.3511 (1.309)			
Fst Genetic Distance * major power dummy					0.1382 (0.625)		
1 if at least one country is a major power					0.3043 (6.256)**		
Fst Gen.Dist * dummy for FST GD > median						0.1325 (0.586)	
Squared Fst genetic distance							-2.9895 (1.915)
# observations	517,251	503,748	517,251	517,251	517,251	517,251	517,251
# of pairs	13,175	12,928	13,175	13,175	13,175	13,175	13,175
Standardized magnitude (%)	-8.521	-9.414	-8.880	-9.014	-8.526	-	-

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 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 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.

Table 9 – Regressions explaining the intensity of conflict
(dependent variable and estimator as described in the second row)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS on conflict intensity variable	**OLS on conflict intensity in subsample with conflict**	**IV on conflict intensity using 1500 GD as IV**	**OLS on number of war casualties**	**OLS on number of war casualties, sample with >0 casualties**	**Probit on War dummy variable (conflict intensity=5)**
Fst genetic distance	**-0.1685** **(8.290)****	**0.2536** **(0.613)**	**-0.2757** **(7.413)****	**-0.1153** **(8.055)****	**-0.9103** **(0.611)**	**-0.0875** **(6.001)****
Log geodesic distance	-0.0330 (4.245)**	0.0032 (0.090)	-0.0263 (3.126)**	-0.0161 (3.549)**	-0.1013 (0.463)	-0.0061 (2.663)**
1 for contiguity	0.2983 (7.693)**	-0.0785 (1.403)	0.2983 (7.696)**	0.0881 (4.982)**	-1.2939 (5.105)**	0.0051 (1.221)
1 if both countries are democracies (polity2>5)	-0.0297 (10.703)**	-0.1377 (1.954)	-0.0317 (10.918)**	-0.0187 (11.204)**	-0.9759 (1.586)	-0.0113 (5.838)**
Number of peaceful years	-0.0008 (9.996)**	-0.0041 (5.990)**	-0.0008 (9.987)**	-0.0004 (9.438)**	-0.0021 (0.641)	-0.0011 (7.190)**
Number of other wars in year t	0.0011 (12.483)**	0.0063 (11.912)**	0.0011 (12.785)**	0.0009 (10.544)**	0.0075 (3.950)**	0.0005 (16.540)**
Dummy for alliance active in year t	-0.0418 (4.955)**	-0.2373 (6.096)**	-0.0414 (4.906)**	-0.0279 (5.994)**	-0.3757 (1.331)	-0.0082 (5.337)**
# observations	517,251	3,844	517,251	516,758	1,531	517,251
# of pairs	13,175	741	13,175	13,175	394	13,175
Adjusted R-squared	0.042	0.135	0.042	0.013	0.252	-
Standardized magnitude (%) [a]	-3.149	1.634	-5.155	-2.722	-2.304	-2.955

Robust t statistics in parentheses (clustering at the country pair level); * significant at 5%; ** significant at 1%.

[a]: In all but columns 6, the standardized magnitude is the standardized beta (effect of a standard deviation change in genetic distance as a percentage of the standard deviation of the dependent variable). For column 6, the standardized magnitude is the effect of a one standard deviation increase in genetic distance as a percentage of the mean probability of war.

Probit marginal effects reported in column (6), where: 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 10 – Post-1950 analysis, controlling for trade variables and income differences (dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Baseline with common sample	**Control for trade terms**	**Additional trade terms**	**Baseline with restricted sample**	**Add per capita income difference**	**Add total income difference**
Fst genetic distance	**-0.5223** **(7.181)****	**-0.3247** **(5.810)****	**-0.3098** **(5.485)****	**-0.3119** **(4.974)****	**-0.3531** **(5.386)****	**-0.3240** **(4.673)****
Log geodesic distance	-0.0154 (2.107)*	0.0047 (0.721)	0.0043 (0.652)	0.0034 (0.509)	0.0018 (0.269)	0.0032 (0.477)
1 for contiguity	0.2528 (6.638)**	0.2186 (7.349)**	0.2134 (7.326)**	0.2313 (7.098)**	0.2762 (7.520)**	0.2326 (7.160)**
Log bilateral openness, t-4		-0.0414 (4.813)**	-0.0395 (4.546)**	-0.0403 (4.327)**	-0.0360 (4.145)**	-0.0400 (4.306)**
Log multilateral openness, t-4		0.0552 (1.993)*	0.0595 (2.191)*	0.0327 (1.032)	0.0158 (0.527)	0.0306 (0.962)
Log distance * log mult. openness		-0.0093 (2.656)**	-0.0098 (2.854)**	-0.0071 (1.789)	-0.0048 (1.289)	-0.0068 (1.700)
Log distance * log bilateral openness		0.0054 (5.095)**	0.0053 (4.883)**	0.0053 (4.627)**	0.0047 (4.398)**	0.0052 (4.490)**
Dummy for zero trade, t-4		-0.0159 (2.382)*	-0.0159 (2.431)*	-0.0137 (1.898)	-0.0127 (1.748)	-0.0149 (2.034)*
Free trade area (full set)			-0.0236 (2.859)**	-0.0226 (2.576)*	-0.0214 (2.310)*	-0.0221 (2.526)*
# of GATT members			-0.0147 (3.768)**	-0.0172 (4.158)**	-0.0162 (3.888)**	-0.0170 (4.134)**
Absolute difference in log per capita income					1.6297 (4.790)**	
Absolute difference in total income						0.1909 (0.773)
# of observations	226,357	226,357	226,357	202,523	202,523	202,523
# of pairs	9,127	9,127	9,127	9,127	9,127	9,127
Standardized magnitude (%)	-6.612	-4.110	-3.921	-4.436	-5.024	-4.609

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.

Table 11: Analysis of UN Vote Correlations, 1946-2000
(Dependent variable: UN Assembly vote correlations; estimator: OLS)

	(1)	(2)	(3)	(4)	(5)
	Univariate	**Add geography controls**	**Full specification**	**1990-2000**	**Excluding abstentions**
Fst genetic distance	**0.2657**	**0.2852**	**0.2923**	**0.1500**	**0.4709**
	(13.356)**	**(14.140)****	**(14.363)****	**(8.033)****	**(15.590)****
Log geodesic distance		0.0240	0.0276	0.0143	0.0435
		(5.764)**	(6.654)**	(4.006)**	(7.406)**
Log absolute difference in longitudes		-0.0199	-0.0186	-0.0081	-0.0282
		(8.052)**	(7.841)**	(3.856)**	(8.251)**
Log absolute difference in latitudes		-0.0387	-0.0370	-0.0355	-0.0694
		(26.599)**	(25.206)**	(24.698)**	(31.779)**
1 for contiguity		0.0535	0.0539	0.0300	0.0806
		(5.271)**	(5.617)**	(2.962)**	(5.835)**
Number of landlocked countries in the pair		0.0171	0.0177	0.0205	0.0177
		(6.661)**	(6.861)**	(9.460)**	(4.561)**
Number of island countries in the pair		0.0298	0.0301	0.0300	0.0408
		(10.593)**	(10.909)**	(11.194)**	(9.562)**
1 if pair shares at least one sea or ocean		-0.0183	-0.0294	-0.0325	-0.0470
		(3.456)**	(5.643)**	(5.344)**	(6.701)**
Log product of land areas in square km		-0.0061	-0.0063	0.0001	-0.0097
		(10.591)**	(10.759)**	(0.234)	(11.683)**
1 if both countries are democracies (polity2>5)			-0.0071	-0.0338	-0.0437
			(2.100)*	(9.939)**	(9.332)**
1 for pairs ever in colonial relationship			-0.1224	-0.1451	-0.2133
			(9.184)**	(10.500)**	(11.464)**
Number of peaceful years			0.0005	-0.0001	0.0007
			(7.853)**	(2.111)*	(7.992)**
Number of other conflicts in year t			0.0013	-0.0004	0.0023
			(32.156)**	(33.041)**	(46.343)**
Dummy for alliance active in year t			0.0614	0.0442	0.1078
			(12.096)**	(6.785)**	(14.887)**
Constant	0.8347	0.9436	0.8311	0.9103	0.6499
	(278.491)**	(31.987)**	(26.381)**	(32.921)**	(14.837)**
Adjusted R^2	0.008	0.074	0.100	0.101	0.147
Standardized beta (%)	9.178	9.853	10.098	6.433	16.266

Robust t statistics in parentheses; * significant at 5%; ** significant at 1%; All regressions run on 385,783 observations from 12,655 country pairs. Note: Ever the same country dummy excluded as regressor as no pair in the sample with UN vote correlations data was made up of countries that were ever a single country.

Figure 1 - Conflict Game in Normal Form

State 1 \ *State* 2	C	N
C	$-\int_{k\in M}\alpha_1(k)(1-\pi)\Delta(k)dk - c_1(C,C)$, $-\int_{k\in M}\alpha_2(k)\pi\Delta(k)dk - c_2(C,C)$	$-\int_{k\in M}\alpha_1(k)(1-\pi-\sigma_1)\Delta(k)dk - c_1(C,N)$, $-\int_{k\in M}\alpha_2(k)(\pi+\sigma_1)\Delta(k)dk - c_2(C,N)$
N	$-\int_{k\in M}\alpha_1(k)(1-\pi+\sigma_2)\Delta(k)dk - c_1(N,C)$, $-\int_{k\in M}\alpha_2(k)(\pi-\sigma_2)\Delta(k)dk - c_2(N,C)$	$-\int_{k\in M}\alpha_1(k)(1-\beta)\Delta(k)dk$, $-\int_{k\in M}\alpha_2(k)\beta\Delta(k)dk$

Payoffs are to state 1 (row), state 2 (column)