

# War and Relatedness\*

Enrico Spolaore  
Tufts University and NBER

Romain Wacziarg  
UCLA, NBER and CEPR

May 2009

## Abstract

We develop a formal theory of interstate conflict in which the degree of genealogical relatedness between populations has a positive effect on their conflict propensities because more closely related populations, on average, tend to interact more and develop more disputes over sets of common rival issues. We examine the empirical relationship between the occurrence of interstate conflicts and the degree of relatedness between countries, showing that populations that are genetically closer 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 measures of trade and democracy.

---

\*Spolaore: Department of Economics, Tufts University, Medford, MA 02155-6722, [enrico.spolaore@tufts.edu](mailto:enrico.spolaore@tufts.edu).  
Wacziarg: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles CA 90095, [wacziarg@ucla.edu](mailto:wacziarg@ucla.edu).  
We are grateful to Klaus Desmet, Rajeev Dehejia, Yannis Ioannides, Michael Klein and Philippe Martin 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.<sup>1</sup> 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.<sup>2</sup> 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.<sup>3</sup> 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 and cultural relatedness between populations - to explore the relationship between kinship and conflict.<sup>4</sup> Genetic distance measures

---

<sup>1</sup>For recent salient examples, see Blainey (1988), Keegan (1984), Ferguson (2006) and Nye (2008).

<sup>2</sup>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 pioneered 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).

<sup>3</sup>The liberal peace view that trade and democracy should reduce the risk of war goes back to Montesquieu (1758) 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), and Levy and Razin (2004).

<sup>4</sup>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 introduced him to the scientific method - hence the fifteen-year project that culminated in the monumental *Study of*

the difference in gene distributions between two populations, where the genes under considerations are neutral: they change randomly and independently of selection pressure, and thus do not affect traits that directly matter for survival and fitness. 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 cultural and historical distance across populations.

This paper's main result is that, surprisingly, genetic distance reduces the risk of conflict. 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. 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 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.<sup>5</sup>

This paper builds on a large and diverse literature. Broad questions about cultural distance, re-

---

*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).

<sup>5</sup>We apologize to Leo Tolstoy for the double plagiarism.

latedness 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: "Ani-mosity 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 culturally and ethnically distant. For example, Bremer (2000, p. 27), referring to evidence from social psychology, wonders whether "cultural differences [...] should lead to misunderstandings, stereotypings, 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 culturally 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 con-tiguity, 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.<sup>6</sup> In their influential study on conflict within states, Fearon and Laitin (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

---

<sup>6</sup>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.

the paramount effect attributed by some scholars to geographic proximity may in part be due to its correlation with cultural and historical relatedness. Once genetic distance is taken into account, geographic variables have smaller effects (although they remain significant). The effect of genetic distance is even higher - and the effects of geography smaller - when we instrument for modern genetic distance using genetic distance between ancestor populations of current countries as of 1500, to account for measurement error and possible endogeneity issues due to post-1500 migrations. The effect of genetic distance is also robust when accounting for other measures of cultural 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 culturally more similar societies also tend to trade more with each other and to share more similar political arrangements (such as democratic regimes). Following this reasoning, the observed low level of conflict might not be the direct effect of trade and democracy, but rather the outcome of deeper cultural similarities (for discussions of this debate see, for example, Schneider, Barbieri and Gleditsch, 2003). In contrast, our estimates provide strong evidence that the premise that closely related populations fight less with each other is incorrect, and hence cannot account for the pacifying effects of bilateral trade and democracy. In sum, our findings validate the liberal view concerning the pacifying effects of trade and democracy.

This paper is the first, to our knowledge, to study the relationship between genetic distance and the likelihood of international conflict and wars. 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.<sup>7</sup> 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 (with important extensions included in Appendix 1). Section 3 introduces our data and methodology. Section 4 discusses the empirical findings. Section 5 concludes.

## 2 A Model of Conflict and Relatedness

In this section we present a stylized model of conflict, which captures the complex interrelations among international disputes, probability of violent conflict and populations' relatedness in a simple way, highlighting the logic of the main mechanisms while maintaining analytical tractability. First (in 2.1), we present a general specification of preferences over sets of issues, and define the concepts of common issues and disputes between states. Second (in 2.2), we model states' choices over war and peace, and derive the equilibrium conditions under which disputes are resolved peacefully or violently. In particular, we show how the probability of violent conflict depends on the extent of common issues. Third (in 2.3), we present a model that links current preferences over issues to the intergenerational transmissions of characteristics, and derive the relation between probability of conflict and degree of genealogical relatedness.

---

<sup>7</sup>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 Survey 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.

## 2.1 Preferences and Common Issues

Consider two political units or states (call them 1 and 2). For simplicity, in the rest of the analysis we will treat a state - or, equivalently, its government - as a unitary agent. Each state cares about a set of issues  $M$ . Each issue  $k \in M$  can take values  $x(k) \in X(k)$ , where  $X(k)$  is a set of real numbers. Each state  $i$ 's preferences over issues can be written as follows:

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

where  $U_i$  is state  $i$ 's utility function,  $x(k)$  is the actual outcome for issue  $k$ ,  $x_i^*(k)$  is state  $i$ 's most preferred outcome for issue  $k$ ,  $\alpha_i(k) \geq 0$  is the weight that state  $i$  attributes to issue  $k$ , and  $c_i$  (net costs from conflict) is equal to zero if all disputes with the other state are solved peacefully, positive otherwise. If  $\alpha_i(k) = 0$ , state  $i$  does not care about issue  $k$ .  $k$  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  $\alpha_2(k) > 0$ . We say that a common issue  $k$  is *disputed* when the two states prefer different outcomes  $x_1^*(k) \neq x_2^*(k)$ . Let  $\Delta(k) \equiv |x_1^*(k) - x_2^*(k)|$  denote the difference between ideal outcomes.

A special but important case is when the issue is about the use of a rival good, where "rival" means that (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. We will call them *common rival issues*. In such cases, it is convenient to define the outcome  $x(k)$  as the extent of state 1's control over the good under dispute, and  $(1 - x_k)$  as the extent of state 2's control, with  $0 \leq x(k) \leq 1$  and  $x_1^*(k) = 1$  and  $x_2^*(k) = 0$ . For example, both states may want to control the same natural or human-made resource (an offshore oil deposit or a religious/cultural center), which both value because they have similar preferences over consumption and/or similar production technologies. This kind of rivalry may also emerge when the inhabitants of the two states interact extensively with each other over a specific set of issues (e.g., migration), and each state wants to impose its exclusive control over those issues. Formally we define as *rival* all issues for which the difference between ideal points is constant and normalized to one  $\Delta(k) = 1$ . This normalization is not essential, but it simplifies notation. What matters is that such rival issues, by definition, imply disagreement between two states caring about the same rival good.

States may also interact with each other over issues that are non-rival in the above-defined sense. For example, both states may care about a common set of international public goods (security of trade routes, regulation of pollution and other environmental externalities), where the

use of the public good by one state's would not reduce the ability of the other state to use the same public good. In this case, disagreement may emerge not about the exclusive use of the good, but about the public good's characteristics (and/or the way to finance it). Within our simplified unidimensional framework, this kind of disagreement can be captured by interpreting  $x(k)$  as the type of international public good  $k$ , and  $x_i^*(k)$  as state  $i$ 's ideal type for that public good. Then, the difference between ideal outcomes may vary across different issues. This does not mean that issues with larger  $\Delta(k)$ 's would necessarily be more relevant for conflict, because the corresponding parameters  $\alpha_k$ 's may vary as well. In general, we will call all cases where  $\Delta(k) \neq 1$  *non-rival* issues.

## 2.2 The Resolution of Disputes

We say that two states face a dispute when one or more common issues are *disputed* - that is, when  $x_1^*(k) \neq x_2^*(k)$ . Disputes between the two states are resolved either peacefully or violently. When a dispute between state 1 and state 2 is resolved peacefully, the outcome for each disputed issue  $k$  is set as follows:

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

where  $\beta$  denotes state 1's bargaining power in a peaceful dispute, with  $0 \leq \beta \leq 1$ . When disputed issues are resolved peacefully, neither state bears conflict costs ( $c_i = 0$ ,  $i = 1, 2$ ). In contrast, if a violent conflict occurs, the winner of the conflict will set all disputed issues according to its preferences. Let  $P$  denote the probability that state 1 will win in a violent conflict (and, therefore, will set each outcome  $k$  equal to its preferred value  $x(k) = x_1^*(k)$ ), while with probability  $1 - P$  state 2 will win and set  $x(k) = x_2^*(k)$  for all  $k \in M$ . 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  $x(k)$ . If the dispute is resolved violently, the expected outcome for each  $x(k)$  is

$$x(k) = Px_1^*(k) + (1 - P)x_2^*(k) \quad (3)$$

Each state can choose whether to "start a conflict" (strategy  $C$ ) or "not to start a conflict" (strategy  $NC$ ). Peace results if and only if both states choose  $NC$ . If both states choose  $NC$ , all issues are settled peacefully, and the states' payoffs are

$$U_1(NC, NC) = - \int_{k \in M} \alpha_1(k)(1 - \beta)\Delta(k)dk \quad (4)$$

$$U_2(NC, NC) = - \int_{k \in M} \alpha_2(k)\beta\Delta(k)dk \quad (5)$$



If both states choose to go to war,  $P = \pi$ , with  $0 \leq \pi \leq 1$ , and  $c_i = \psi_i > 0$ , and the states' payoffs are

$$U_1(C, C) = - \int_{k \in M} \alpha_1(k)(1 - \pi)\Delta(k)dk - \psi_1 \quad (6)$$

$$U_2(C, C) = - \int_{k \in M} \alpha_2(k)\pi\Delta(k)dk - \psi_2 \quad (7)$$

If state 1 chooses  $C$  while state 2 chooses  $NC$ , we assume  $P = \pi + \sigma_1$  with  $0 < \sigma_1 \leq (1 - \pi)$ .  $\sigma_1$  captures the increased probability of winning that results from a "first-mover's advantage". The costs of conflict will be  $c_1 = \phi_1 \geq 0$  and  $c_2 = \phi'_2 \geq \psi_2$ , and the payoffs are:<sup>8</sup>

$$U_1(C, NC) = - \int_{k \in M} \alpha_1(k)(1 - \pi - \sigma_1)\Delta(k)dk - \phi_1 \quad (8)$$

$$U_2(C, NC) = - \int_{k \in M} \alpha_2(k)(\pi + \sigma_1)\Delta(k)dk - \phi'_2 \quad (9)$$

Analogous equations hold for  $U_1(NC, C)$  and  $U_2(NC, C)$ .<sup>9</sup>

The peaceful outcome  $(NC, NC)$  is a Nash equilibrium if and only if:<sup>10</sup>

$$U_1(NC, NC) \geq U_1(C, NC) \quad (10)$$

and:

$$U_2(NC, NC) \geq U_2(NC, C) \quad (11)$$

Hence, the conditions for peace to be an equilibrium are:

$$(\pi - \beta + \sigma_1) \int_{k \in M} \alpha_1(k)\Delta(k)dk \leq \phi_1 \quad (12)$$

$$(\beta - \pi + \sigma_2) \int_{k \in M} \alpha_2(k)\Delta(k)dk \leq \phi_2 \quad (13)$$

These two conditions (12) and (13) can be written more compactly by making two simplifying assumptions:

---

<sup>8</sup>This assumption means 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 "voluntarily" (i.e., if it had selected C rather than NC).

<sup>9</sup>When state 1 chooses  $NC$  and state 2 chooses  $C$ ,  $P = \pi - \sigma_2$ , with  $0 < \sigma_2 \leq \pi$ ,  $c_1 = \phi'_1$  and  $c_2 = \phi_2$ , and payoffs are  $U_1(NC, C) = - \sum_{k=1}^M \alpha_{2k}|(1 - \pi + \sigma_2)(x_{1k}^* - x_{2k}^*)| - \phi'_1$   $U_2(NC, C) = - \sum_{k=1}^M \alpha_{2k}|(\pi - \sigma_2)(x_{1k}^* - x_{2k}^*)| - \phi_2$

<sup>10</sup>This stylized model of war is consistent with Fearon's (1995) discussion of violent conflict as emerging from an inability to commit to a Pareto-superior outcome. In our framework both states would be better off if each could commit to play NC, but they can do that credibly if and only if  $(NC, NC)$  is a Nash equilibrium.

(i)  $\beta = \pi$  - which means that the bargaining power of each state in a peaceful negotiation depends on its probability of winning a conflict should the states go to war. This is motivated by the reasonable hypothesis that peaceful negotiations typically take place under the shadow of war, and that a state's bargaining power depends on its strength should negotiations break up - a common assumption in the literature (for example, see Alesina and Spolaore, 2005).

(ii)  $\Delta(k) = 1$  for all  $k$  (that is, assuming that all disputes are about rival issues).

Under the above assumptions, peace  $(NC, NC)$  is an equilibrium if and only if:

$$\max_{i=1,2} \left\{ \int_{k \in M} \alpha_i(k) dk - \frac{\phi_i}{\sigma_i} \right\} \leq 0 \quad (14)$$

In contrast, if  $\max_i \left\{ \int_{k \in M} \alpha_i(k) dk - \frac{\phi_i}{\sigma_i} \right\} > 0$ , conflict  $(C, C)$  is the unique Nash equilibrium.<sup>11</sup>

The intuition for the above conditions is straightforward: conflict is more likely: (1) the higher is  $\int_{k \in M} \alpha_i(k) dk$  - that is, the larger the set of common issues under dispute and the extent the two states care about them; (2) the lower is the cost of starting a war  $\phi_i$ ; and (3) the higher is the temptation to start a war  $\sigma_i$ .

### 2.3 Relatedness and the Probability of Conflict

The above analysis relates the conditions for conflict to the set of common issues under dispute. However, so far we have taken such issues as given. In what follows we will go a step further, and model the set of common issues as the outcome of long-term connections between the populations of the two states. The general idea is that long-term relatedness will affect what issues a state will (a) care about, and (b) interact about with other states. If there is some long-term persistence of preferences and characteristics across time, and current populations inherit such preferences and characteristics with variation from their ancestors, on average populations that are more closely related genealogically will be more likely to share a large range of common issues. As we will see, such common-issues effect implies a *higher* likelihood of conflict across more *closely* related

---

<sup>11</sup>Even when  $\max \sum_{k=1}^M \alpha_{ik} - \omega_i \leq 0$ ,  $(C, C)$  remains a Nash equilibrium, because if any state plays C, the other state is better off to play C rather than NC, given that  $\sigma_i > 0$  and  $\phi'_i \geq \psi_i$ . However, under assumption (i), both states are better off in a peaceful negotiation  $(NC, NC)$  than in the case of conflict  $(C, C)$ , because the outcomes  $x$ 's are identical in the two cases, but the two players do not bear the costs from conflict in  $(NC, NC)$ . Consequently, if  $(NC, NC)$  is a Nash equilibrium, it is the unique coalition-proof Nash equilibrium, as defined by Bernheim, Peleg and Whinston (1987). We return to this point in 2.3.

populations, other things being equal. In Appendix 1 we model other effects (extent of disagreement and coordination failure) which, in principle, might serve to increase or reduce the common-issues effect.

To simplify notation, in the rest of the analysis we will assume that (a) peaceful bargaining takes place under the shadow of war:  $\beta = \pi$ , (b) all states face the same ratio  $\phi_i/\sigma_i \equiv \omega$ , and (c)  $M$  is the set of real numbers. Moreover, we will also assume that all common issues are rival - that is,  $\Delta(k) = 1$  (the assumption is relaxed in Appendix 1).

Assume that each state  $i$  cares about a mass  $R_i$  of issues. A given state  $i$  will care about all issues between point  $a_i$  and point  $b_i > a_i$ , with  $b_i - a_i = R_i$ , but will not care about issues outside that range. In addition, assume that all relevant issues receive equal weight  $\alpha > 0$  - that is,  $\alpha_i(k) = \alpha > 0$  if and only if  $a \leq k \leq b$ , while  $\alpha_i(k) = 0$  otherwise. Finally, assume that  $R_i = R$  for all states. These assumptions allow us to characterize the set of issues that state  $i$  cares about by a single real number  $v_i$ , which we can interpret as that state's fundamental characteristics or type affecting what issues the state cares about. To fix ideas, let  $v_i$  be the mid-point in state  $i$ 's set of relevant issues:

$$v_i = a_i + \frac{R}{2} = b_i - \frac{R}{2} \quad (15)$$

Therefore, a state with characteristics  $v_i$  will have the following preferences:<sup>12</sup>

$$U_i = - \int_{v_i - \frac{R}{2}}^{v_i + \frac{R}{2}} \alpha dk - c_i \quad (16)$$

Let  $V(i, j) \equiv |v_i - v_j|$  denote the distance between state  $i$  and state  $j$  in their fundamental characteristics. Violent conflict between a state of type  $v_i$  and a state of type  $v_j$  will *not* occur if  $V(i, j) \equiv |v_i - v_j| > R$ . This captures the obvious but important point that two states which are very far in the set of issues they care about will have no reason to fight. In contrast, if  $V(i, j) < R$  there will be a range of common issues between the two states. If  $v_i \leq v_j$ , the common range of issues will include all issues between  $v_j - \frac{R}{2}$  and  $v_i + \frac{R}{2}$ , and violent conflict will occur whenever  $\int_{v_j - \frac{R}{2}}^{v_i + \frac{R}{2}} \alpha dk > \omega$ . By the same token, if  $v_i \geq v_j$ , violent conflict will occur whenever  $\int_{v_i - \frac{R}{2}}^{v_j + \frac{R}{2}} \alpha dk > \omega$ .

---

<sup>12</sup>In this simplified analysis we assume that each political unit is a unified agent, formed by one population with homogeneous characteristics  $v_j$ . In principle, two states can be of the same type - that is, they may care about the identical set of issues. In this basic framework we abstract from the possibility that states may include mixed populations with different preferences over issues (population heterogeneity within states will be taken into account in the empirical analysis).

In sum, violent conflict will be the unique equilibrium if and only if:

$$\alpha[R - V(i, j)] > \omega \tag{17}$$

This simple inequality illustrates an important result: for a given range  $R$  of common issues, populations which are more *distant* in preferences over relevant issues - i.e., populations with a larger  $V(i, j) \equiv |v_i - v_j|$  - are *less* likely to go to war with each other. In particular, if  $\omega$  is a random variable distributed uniformly between 0 and  $\bar{\omega}$ , we have the following:

**Proposition 1**

*The probability of conflict between two states  $i$  and  $j$  is given by*

$$\text{Prob}(\text{Conflict}) = \frac{\alpha}{\bar{\omega}}[R - V(i, j)] \tag{18}$$

*That is, conflict is increasing in the extent states care about specific common issues ( $\alpha$ ) and the range of common issues each state cares about ( $R$ ), decreasing in the (expected) costs of fighting divided by the temptation to start a fight ( $\bar{\omega}$ ), and decreasing in the distance between the two states' fundamental characteristics  $V(i, j) \equiv |v_i - v_j|$ .*

Testing for such relationship between probability of militarized conflict and distance in fundamental characteristics is at the center of our empirical analysis. In order to link our theoretical analysis with the empirics, a further step is needed. In what follows we are going to derive the relationship between the probability of conflict and actual measures of long-term relatedness (genetic distance). To do so we take advantage of the fact that characteristics determining preferences over common issues are transmitted intergenerationally across populations (with variation), and therefore, on average, tend to be a function of populations' long-term genealogical relatedness. As we will see, this implies that, wars will occur more frequently across states inhabited by populations that are genealogically more closely related, because those populations will be more likely to care about the same issues.

This can be easily illustrated by considering a simple model of the transmission of characteristics  $v_{it}$ . Assume that in period  $t$  a population  $i$  inherits its characteristics  $v_{it}$  from an ancestor population with characteristics  $v_{it-1}$ , with some random variation, denoted by a random shock  $\varepsilon_{it}$ :

$$v_{it} = v_{it-1} + \varepsilon_{it} \tag{19}$$

For example, consider the very simple case in which  $\varepsilon_{it}$  follows a simple random walk, taking value  $\varepsilon > 0$  with probability 1/2 and value  $-\varepsilon$  with probability 1/2, and the shocks are independently

distributed across different populations. After one period, two populations that share the same ancestor population will be at distance  $V(i, j) = 0$  with probability  $1/2$  and distance  $V(i, j) = 2\varepsilon$  with probability  $1/2$ . hence, their expected distance will be  $E\{V(i, j)\} = \varepsilon$ . By contrast, two populations sharing a two-period-old last common ancestor population will be at a higher expected distance  $E\{V(i, j)\} = 3\varepsilon/2$ .

In general, the farther in time we have to go back to find the last common ancestors between two populations, the larger the expected distance in inherited characteristics, and hence the smaller the set of common issues and interactions between the two populations within this model. An immediate implications of this model, therefore, is that genetic distance  $g(i, j)$  between the population of state  $i$  and the population of state  $j$ , measured by the time since the two populations shared their last common ancestor, will be inversely related to their expected probability of conflict. In the empirical analysis, we measure relatedness by  $F_{ST}$  genetic distance, a measure that is approximately linear in the time since two populations shared their last common ancestors.

If the parameters are such that  $V(i, j) < R$  for all possible realizations of the shocks, we have the following:<sup>13</sup>

**Corollary 1**

The *expected probability of conflict between two states  $i$  and  $j$*  is given by:

$$E[\text{Prob}(\text{Conflict})] = \frac{\alpha}{\omega}[R - g(i, j)] \tag{20}$$

where  $g(i, j)$  measures their respective genetic distance.

Corollary 1 is our central theoretical result, which we are going to test directly in the empirical section. Even though this result is obtained under a series of simplifications and abstractions, it highlights the more general logic of the relation among common issues, interactions across states, long-term relatedness, and probability of conflict. As we will see, it is consistent with the empirical evidence.

One could think of our basic results as the reduced forms of more detailed and micro-founded settings in which specific interactions and common issues emerge from more complex dynamic processes and decisions. For example, societies with more similar long-term characteristics might

---

<sup>13</sup>The expected effect of genetic distance  $g(i, j)$  on the probability of conflict will be even bigger if we consider the case when  $V(i, j) > R$ , as expected conflict would then be zero, on average, for larger  $g(i, j)$ .

endogenously end up with more similar production systems and/or consumption patterns, which may induce them to compete over a similar set of resources. Another (non-mutually exclusive) channel would emerge if genealogically more similar populations face lower fixed costs to interacting with each other, and therefore have more incentives and opportunities to interact over all sorts of common issues, multiplying the likelihood that some of those issues will be disputed. These interpretations are consistent with our simplified framework, as they predict a negative relationship between genetic distance and the probability of conflict.

### 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 (in an ambiguous direction) and may affect the likelihood of reaching peaceful conflict resolution by facilitating coordination (corollaries 2 and 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](http://www.correlatesofwar.org)).<sup>14</sup> 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 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

---

<sup>14</sup>See also Jones et. al. (1996) and Faten et al. (2004).

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 mostly,  $F_{ST}$  genetic distance, captures the length of time since two populations became separated from each other. When two populations split apart, random genetic mutations result in genetic differentiation over time. The longer the separation time, the greater the genetic distance computed from a set of neutral genes. In other words,  $F_{ST}$  genetic distance is a direct measure of genealogical relatedness, resulting from a molecular clock. The specific source for our data is Cavalli-Sforza et al. (1994), pp. 75-76.<sup>15</sup>

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  $m = 1 \dots M$  and country  $j$  is composed of populations  $n = 1 \dots N$ . Denote by  $s_{1m}$  the share of population  $m$  in country  $i$

---

<sup>15</sup>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.

(similarly for country  $j$ ) and  $d_{mn}$  the genetic distance between populations  $m$  and  $n$ . The weighted  $F_{ST}$  genetic distance between countries  $i$  and  $j$  is then:

$$FST_{ij}^W = \sum_{m=1}^M \sum_{n=1}^N (s_{im} \times s_{jn} \times d_{mn}) \quad (21)$$

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 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.<sup>16</sup>

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 the match 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 Aborigenes. Genetic distance between countries using the 1500 match can be used as an instrument for current genetic distance.<sup>17</sup>

### 3.3 Summary Statistics

Table 1 and 2 provide summary statistics that can help to give a sense of the data and provide clues concerning the relationship between conflict and relatedness.<sup>18</sup> The baseline sample is an

---

<sup>16</sup>All our results are robust to using genetic distance between plurality groups rather than weighted genetic distance. The corresponding estimates are available upon request.

<sup>17</sup>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.

<sup>18</sup>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.



unbalanced panel of 517,251 observations covering 13,575 country pairs, based on 176 underlying countries, from 1816 to 2001. 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 to our surprise, a relatively small portion of full fledged-wars occur between contiguous countries (18.2%), and the mean geodesic distance separating countries at war is relatively high (5,562 km).

Table 2 shows the conditional frequency of both wars and conflicts. Wars are a relatively rare occurrence, 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. 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) of regressing a binary indicator of

interstate conflict on a set of bilateral determinants. The baseline regression equation is:

$$C_{ijt} = \beta X_{ijt} + \gamma FST_{ij}^W + \varepsilon_{ijt} \quad (22)$$

The vector  $X_{ijt}$  contains a series of controls such as a contiguity dummy, log geodesic distance, log longitudinal and latitudinal distance, several other indicators of geographic isolation, as well as a set of dummy variables representing whether both countries in the pair are democracies, whether they were ever in a colonial relationship, whether they belong to an active military alliance, among other controls. The choice of controls follows the existing literature closely, 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). Equation (22) is estimated using probit, clustering standard errors at the country-pair level. Throughout, we report marginal effects evaluated at the mean of the independent variables, providing a quantitative assessment of the magnitude of the effects. 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 Baseline Estimates

Table 3 presents baseline estimates of the coefficients in equation (22). 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 (0.066) associated with a 0.492 decline in the percentage probability of conflict (the mean of this variable, again, is 0.721). 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. These measures usually bear the expected signs, and their inclusion greatly reduces the effect of genetic distance.<sup>19</sup> However, this effect remains negative and

---

<sup>19</sup>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

highly significant statistically. Its magnitude is still substantial - a one standard deviation shift in genetic distance is associated with a reduction in the probability of conflict of 12.15% of that variable's mean.

Several other factors have been proposed as correlates of war. Chief among them is 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 controls such as whether countries in a pair ever had colonial ties, the number of peaceful years 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. Once all these controls are included, the coefficient on genetic distance falls further, but remains negative and significant at the 1% level. The effect of a one standard deviation shift in genetic distance, with the full set of controls, remains equal to 8.52% of the mean probability of conflict. We continue to condition on this full set of controls in all the regressions that follow.

None of these observations change very much when using a logit estimator rather than a probit estimator (column 4). We continue to use a probit estimator in the rest of this paper. Finally, in column 5, we instrument for genetic distance using genetic distance between populations as they were in 1500. The results are very close to those previously reported, but the effect of genetic distance rises by over 50% relative to the estimates of column 3, suggesting that the latter understated the effect. It is likely that the higher effect of genetic distance under IV reflects the fact that measurement error is less prevalent, since arguments about reverse causality or omitted variables bias would suggest that instrumenting should reduce the effect of genetic distance. To adopt a conservative approach, we refrain from instrumenting for genetic distance in the bulk of our analysis, keeping in mind that our reported effect is likely an understatement of the true magnitude.

## 4.2 Estimates Across Time and Space

To examine if specific periods or regions account for the finding of a negative effect of relatedness on conflict, we broke down the sample by time period and region. Results are presented in Tables 4 and

---

of relatedness from past empirical specifications seeking to explain conflict likely led to overstating the quantitative impact of geographic factors.

5. We find that results are remarkably robust within regions and periods. Table 4 shows that 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.<sup>20</sup> The coefficient is negative whatever the subperiod under consideration.

Turning to the regional breakdown in Table 5, we again uncover a negative effect of genetic distance on conflict whatever the region under consideration. Column (2) starts by including a dummy variable taking on a value of one if both countries in a pair are part of the same continent (continents are defined as Africa, Americas, Asia, Europe, and Oceania). The concern is that conflicts occur predominantly among countries located on the same continent (this was the case for 2,086 out of a total of 3,728 conflicts in our baseline sample), and that populations located on the same continent tend to be genetically close (the mean of  $F_{ST}$  genetic distance for pairs on the same continent is 0.066 versus a sample-wide mean of 0.102). However, the inclusion of the same continent dummy hardly changes the coefficient on genetic distance at all.

Column (3) presents results for Europe. For this continent, we observe a separate matrix of  $F_{ST}$  genetic distances, available for almost all the countries in Europe.<sup>21</sup> Despite the paucity of observations (only 291 country pairs), the effect of genetic distance remains negative and significant at the 5% level. A one standard deviation change in genetic distance reduces the probability of conflict by 12.531% of its mean, a magnitude slightly larger than, but roughly in line with, the results found in the World sample. Columns (4) through (6) provide estimates for Asia, Africa and the Americas (there were no conflicts within Oceania in our baseline sample, so this category is missing). The coefficient on genetic distance is consistently negative, and significant at the 10%

---

<sup>20</sup>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.

<sup>21</sup>Details concerning the  $F_{ST}$  genetic distance matrix for the European continent can be found in Spolaore and Wacziarg (2009). There are only 5 distinct European populations in the worldwide matrix, so estimates using the European matrix, where there are 26 distinct genetic groups, are likely to be much more reliable.

level for Asia and Africa, but small and insignificant for the Americas.<sup>22</sup> Overall, the regional breakdown suggests that the negative effect of relatedness on war is remarkably consistent across space, the results within Europe, where genetic distance is small, being particularly striking.

### 4.3 Adding Linguistic and Religious Distance

While genetic distance is a precise and continuous measure of the degree of relatedness between populations and countries, alternative 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).<sup>23</sup> Religious beliefs, also transmitted intergenerationally, are one type of difference in human traits that can lead to 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.<sup>24</sup>

Prior to showing the results, we briefly discuss 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

---

<sup>22</sup>The number of intracontinental interstate conflicts experienced by these continents were 787 (Asia), 252 (Africa) and 433 (Americas).

<sup>23</sup>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 results of the edict of foreign rulers, as happened when the Magyar rulers imposed their language on the Hungarian population. Other salient examples include countries were 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 in-depth discussion of these points.

<sup>24</sup>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.

15 nested classifications, so two countries with populations speaking the same language will share 15 common nodes).<sup>25</sup> Using data on the distribution of each linguistic group within and across 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 21).<sup>26</sup> Following Fearon (2003), we transformed these measures so that they reflect linguistic distance (LD) rather than similarity, and are bounded by 0 and 1:

$$LD = \sqrt{\frac{(15 - \# \text{ Common Nodes})}{15}} \quad (23)$$

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 Meham, 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 Meham, 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 (23), again computing plurality and weighted distances separately.

Table 6 presents estimates of the effect of genetic distance on the propensity for interstate conflict when linguistic and religious distance are included. Since the use of these variables constrains the sample (a loss of some 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

---

<sup>25</sup>As an alternative, we used a separate measure of linguistic distance, based on lexicostatistics, from Dyen, Kruskal and Black (1992). This is a more continuous measure than the one based on common nodes, but it is only available for countries speaking Indo-European languages. It captures the number of common meanings, out of a list of 200, that are conveyed using "cognate" or related words. Summing over the 200 meanings, a measure of linguistic distance is the percentage of non-cognate words. Using the expected (weighted) measure of cognate distance led to effects of genetic distance very similar to those obtained when controlling for the Fearon measure, albeit on a much smaller sample of countries speaking Indo-European languages. These results are available upon request.

<sup>26</sup>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 weighted measure did not make any difference for our results. As we did for genetic distance, we focus on weighted measures.

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 significance level drops to 13% when including linguistic distance along with religious distance.<sup>27</sup> 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.

#### 4.4 Nonlinearities and Determinants of Conflict Intensity

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

In columns (2) through (4) of Table 7 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 sum, we find no evidence that the effect of genetic distance depends on some characteristic of the pairs, or that it is nonlinear.

Table 8 seeks to explain the intensity of militarized conflict as opposed to its incidence only. To do so, we modified the dependent variable in several ways. Column (1) simply uses the measure of the intensity of conflict from the Correlates of War dataset, rather than the binary transform of this

---

<sup>27</sup>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).

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), which limits 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 Table 3, 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 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 2.956% 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%).

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

#### **4.5 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 (1758) and Kant (1795), holds that extensive bilateral commercial links between countries reduces the probability of conflict, essentially by raising its cost, since valuable trade links would be lost in a militarized conflict. In an important 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.<sup>28</sup> 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 9. In column (1), we replicate the baseline specification for the smaller sample covering 1950-2000. We are able to exactly 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 findings 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 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 variable 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

---

<sup>28</sup>The data was obtained from <http://team.univ-paris1.fr/teamperso/mayer/data/data.htm>

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).<sup>29</sup> The coefficient on income difference is statistically indistinguishable from zero, and its inclusion does not affect the coefficient on genetic distance.

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

## 5 Conclusion

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 and wars 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. 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 cultural and historical relatedness. As we have discussed in the introduction, our results provide strong evidence against the primordialist view that cultural and ethnic dissimilarity should breed war and plunder.

More broadly, this paper is part of a growing literature in political economy that focuses on the effects of long-term cultural and historical variables on political, economic and institutional outcomes, both theoretically and empirically. For example, it would be interesting to link our approach to the extensive literature on ethnic fractionalization and polarization within countries (e.g., 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 findings

---

<sup>29</sup>The source for the income data is the Penn World Tables, version 6.1 (Heston, Summers and Aten, 2002).

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.<sup>30</sup>

## References

- Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg, "Fractionalization," *Journal of Economic Growth*, 8 (2003), 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), June: 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).
- Bremer, S. A. (1992), "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War 1816-1965," *Journal of Conflict Resolution*, 36 (2): 309-341.

---

<sup>30</sup>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).

- 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, November.
- Dyen, I., J. B. Kruskal and P. Black (1992), "An Indoeuropean 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.
- Gleditsch, N. P. and J. D. Singer (1975), "Distance and International War, 1816-1965," in *Proceedings of the International Peace Research Association (IPRA) Fifth General Conference* (Oslo: IPRA): 481-506.

- Henderson, E. A. (1997), "Culture or Contiguity. Ethnic Conflict, the Similarities of States, and the Onset of War, 1820-1989, *Journal of Conflict Resolution*, 41, 5, 649-668.
- Hess, G. D., Orphanides, A., (1995) "War Politics: an Economic Rational-Voter Framework," *American Economic Review* 85 (4), 828-846. (September).
- Hess, G. D., Orphanides, A., (2001) "War and Democracy," *Journal of Political Economy*, 776-810. (August).
- Heston, Alan, Robert Summers, and Bettina 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
- 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), September: 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, Immanuel (1795), *Toward Perpetual Peace* (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. Russett (1993), "Normative and Structural Causes of the Democratic Peace, 1946-1986," *American Political Science Review*, 89.
- Martin, P., T. Mayer and M. Thoenig (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. Quinn, James Fearon and David 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 (1758), *The Spirit of the Laws* (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 University Press, Princeton, NJ
- 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 Organization*. 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*, September-October, 3-50.
- Spolaore, E. (2008) "Civil Conflict and Secessions," *Economics of Governance*, January, 9 (1): 45-63.
- Spolaore E. and R. Wacziarg (2009), "The Diffusion of Development," *Quarterly Journal of Economics*, 124 (2), May.

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

### Extensions to the Model: Extent of Disagreement and Coordination Failure

In this appendix, we extend our basic framework to include two forces which, in principle, could change the direction of the effect of relatedness on conflict. They have to do with (1) the extent of disagreement over non-rival issues, and (2) the possibility of coordination failure.

#### A1.1 *Extent of Disagreement Over Non-Rival Issues*

When disputes are not fully rival, the extent of disagreement, measured by  $\Delta_{ij}(k) = |x_i^*(k) - x_j^*(k)|$ , may vary, depending on the characteristics of the two states, including their genetic distance. For example, states might agree to exercise *joint* control over an issue (say, provide some joint public good), and only disagree about some specific aspect of the issue (say, location or characteristics of the organization). In such context, it is possible that culturally closer population may face less *disagreement* over such non-rival common issues, and therefore face lower  $\Delta_{ij}(k)$ , which, other things being equal, would *reduce* the probability of conflict, as suggested by equations (14) and (15) However, such effect would only hold for a *given set* of non-rival common issues. It would still be true that more closely related populations, on average, would share a *larger* set of common issues (both rival and non-rival). Such common-issue effect, as we have seen in Section 2, will go in the direction of *increasing* conflict. The net effect of genetic distance on actual conflict would then depend on whether the common-issue effect highlighted in our basic model trumps any countervailing effect from lower disagreement over non-rival issues. It is also worth noting that,

in principle, the relationship between extent of disagreement over non-rival issues and distance in fundamental characteristics may go in the opposite direction, with closer populations disagreeing more over the choice of public goods (say, because a long history of common disputes over rival and non-rival issues might have amplified their differences). In this latter case, such disagreement effect would amplify the results in Propositions 1 and 2 (i.e., it would strengthen the negative relation between probability of conflict and relatedness).

A simple formalization capturing such additional force can be sketched as follows. Suppose that for any set of issues between any two points  $a$  and  $b$  on the real line, a fraction  $\rho$  is rival and a fraction  $(1 - \rho)$  is non-rival. Moreover, assume that for all non-rival common issues between state  $i$  and state  $j$ , their disagreement  $\Delta(i, j)$  is increasing in their distance  $V(i, j) \equiv |v_i - v_j|$ :

$$\Delta(i, j) = \Delta_0 + \delta V(i, j) \tag{A1}$$

where  $\Delta_0 \geq 0$  and  $\delta$  is a parameter measuring the relation between distance  $V(i, j)$  and disagreement  $\Delta(i, j)$ . The other assumptions of our model are maintained. In particular, it is still the case that 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 will be the only equilibrium if and only if  $\alpha[\rho + (1 - \rho)\Delta][R - V(i, j)] > \omega$ . To simplify notation, we assume that all relevant issues share the same  $\alpha$ . The results would not change if we assume that relevant rival issues enter into a state's utility with parameter  $\alpha_r > 0$  while non-rival issues enter with parameter  $\alpha_{nr} > 0$ . Assuming again that  $\omega$  is a random variable distributed uniformly between 0 and  $\bar{\omega}$ , we now have:

**Proposition 2**

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

$$\text{Prob}(\text{Conflict}) = \frac{\alpha}{\bar{\omega}} \{[\rho + (1 - \rho)\Delta_0]R - [\rho + (1 - \rho)(\Delta_0 - \delta)]V(i, j) - (1 - \rho)\delta[V(i, j)]^2\} \tag{A2}$$

*A negative relationship between probability of conflict and distance  $V(i, j)$  holds if and only if the effect of distance  $V(i, j)$  on disagreement  $\Delta(i, j)$  is small enough (low  $\delta$ ), that is,  $\frac{d\text{Prob}(\text{Conflict})}{dV(i, j)} < 0$  if and only if:*

$$\delta < \frac{\rho + (1 - \rho)\Delta_0}{(1 - \rho)[1 - 2V(i, j)]} \tag{A3}$$

The above inequality is always satisfied if  $\delta \leq 0$ . If  $\delta > 0$ , the inequality is more easily satisfied (a) the larger the fraction of rival issues  $\rho$ ; (b) the larger the extent of disagreement which is independent of distance  $\Delta_0$ ; and (c) the larger the distance between states  $V(i, j)$ . Since expected



$V(i, j)$  is a linear function of  $g(i, j)$ , Proposition 2 can be restated as Corollary 2 in terms of expected conflict and genetic distance, in analogy to Proposition 1 and its Corollary 1:

**Corollary 2**

*Expected conflict is decreasing in genetic distance if and only if inequality (A3) holds.*

Corollary 2 implies that an inverse relationship between conflict and genetic distance is consistent with a small or negative effect of distance  $V(i, j)$  on disagreement, and/or with a predominance of rival issues in international disputes.

*A1.2 Coordination failure*

The equilibrium conditions for peace - derived in Section 2 - ensure that (NC, NC) is the unique coalition-proof Nash equilibrium. However, as we have noted, (C, C) is always a Nash equilibrium, because if any state plays C, the other state is always better off if it plays C rather than NC. Nonetheless, both states would be better off with a peaceful negotiation than with violent conflict (C,C), because of the costs of war and violence, which are not borne in a peaceful outcome. Consequently, if (NC, NC) 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 equilibrium, and end up in the inferior conflict equilibrium, even when the conditions for a peaceful equilibrium are satisfied? The question then is whether such coordination failure may be more likely across populations that are genealogically more distant, because those populations, by having more different norms, habits, languages etc., may find it more difficult to communicate and coordinate on a better outcome. If this is indeed the case, such "coordination failure effect" would reduce the negative correlation between historical distance and probability of conflict, all other things being equal. In contrast, if, for any reason, coordination failure were more likely between more closely related populations, the effect of relatedness on conflict would be strengthened.

A formalization of these ideas can be 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 more likely if the states are distant in their preferences and characteristics, measured by  $v_i$  and  $v_j$ . Specifically, assume that:

$$\chi(i, j) = \chi_0 + \theta V(i, j) \tag{A4}$$

with  $\chi_0 \geq 0$  and  $\theta$  is a parameter measuring the relation between distance  $V(i, j)$  and probability of coordination failure  $\chi(i, j)$ . We 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  $\omega$  is a random variable distributed uniformly between 0 and  $\bar{\omega}$ , we have:

**Proposition 3**

*The probability of conflict between the two states when all common issues are rival ( $\rho = 1$ ) can be written as:*

$$\begin{aligned} \text{Prob}\{Conflict\} &= \chi(i, j) \text{Prob}[\alpha[R - V(i, j)] \leq \omega] + \text{Prob}[\alpha[R - V(i, j)] > \omega] = \quad A5 \\ &= \chi(i, j) + [1 - \chi(i, j)] \frac{\alpha}{\bar{\omega}} [R - V(i, j)] \end{aligned}$$

*The probability of conflict is decreasing in  $V(i, j)$  if the effect of distance  $V(i, j)$  on the probability of coordination failure  $\chi(i, j)$  is small enough (low  $\theta$ ), that is,  $\frac{d\text{Prob}(Conflict)}{dV(i, j)} < 0$  if and only if:*

$$\theta < \frac{1 - \chi_0}{\frac{\bar{\omega}}{\alpha} - R - 2V(i, j)} \quad (A6)$$

*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  $\theta$ , it is more easily satisfied for smaller  $\chi_0$ , larger  $\alpha$  and  $R$ , and larger  $V(i, j)$ .

Proposition 3 (like the two previous Propositions) has direct implications for the relation between probability of conflict and genetic distance, which can be summarized as:

**Corollary 3**

*Expected conflict is decreasing in genetic distance if and only if inequality (A6) holds (or, for the more general case  $\rho \leq 1$ , if the inequality (A9) below holds).*

Corollary 3 implies that, if observed violent conflict partly stems from coordination failures, an inverse relationship between conflict and genetic distance is consistent with a small or negative effect of distance on such coordination failures.

**Generalization of Proposition 3 for  $\rho \leq 1$**

For the more general case  $\rho \leq 1$ , we have:

$$\text{Prob}\{Conflict\} = \chi(i, j) + [1 - \chi(i, j)] \text{Prob}\{C, C \text{ Unique Eq}\} \quad (A7)$$

where:

$$\text{Prob}\{C, C \text{ Unique Eq}\} = \frac{\alpha}{\bar{\omega}} \{[\rho + (1 - \rho)\Delta_0]R - [\rho + (1 - \rho)(\Delta_0 - \delta)]V(i, j) - (1 - \rho)\delta[V(i, j)]^2\} \quad (\text{A8})$$

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 - i.e.,  $\frac{d \text{Prob}(\text{Conflict})}{dV(i, j)} < 0$ , if and only if:*

$$\theta < \frac{(1 - \chi_0) \{ \rho + (1 - \rho)(\Delta_0 - \delta) + 2(1 - \rho)\delta V(i, j) \}}{\frac{\bar{\omega}}{\alpha} - [\rho + (1 - \rho)\Delta_0]R - 2[\rho + (1 - \rho)(\Delta_0 - \delta)]V(i, j) - 3(1 - \rho)\delta[V(i, j)]^2} \quad (\text{A9})$$

which reduces to the condition  $\theta < \frac{1 - \chi_0}{\frac{\bar{\omega}}{\alpha} - R - 2V(i, j)}$  for  $\rho = 1$ .

## 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, weighted	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
1 - % cognate measure of linguistic distance, weighted	76,330	0.531	0.292	0	0.918

### 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, weighted	-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)
1 - % cognate, weighted	0.003 (76,330)	0.008* (76,330)	-0.104* (76,330)	0.160* (76,330)	-0.177* (76,330)	0.632* (67,657)	0.823* (67,657)

(# of observations in parentheses)

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

	<b>Fst Genetic Distance, weighted</b>	<b>Geodesic Distance, km</b>	<b>Contiguity dummy</b>
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)**

<b>Conditioning statement:</b>	<b>Bottom decile of genetic distance</b>	<b>Bottom quartile of genetic distance</b>	<b>Third quartile of genetic distance</b>	<b>Second quartile of genetic distance</b>	<b>Top quartile of genetic distance*</b>	<b>Total</b>
	<b>Hostility level = 5 (War)</b>					
None	277	487	178	301	44	1,010
Common sea / ocean = 0	170	329	129	269	44	771
Contiguity = 0	175	368	123	291	44	826
Distance > 1000 km	163	349	155	289	44	837
	<b>Hostility Level &gt; 3 (Conflict)</b>					
None	1,076	1,937	940	717	134	3,728
Common sea / ocean = 0	582	1,128	569	553	114	2,364
Contiguity = 0	537	1,202	520	616	119	2,457
Distance > 1000 km	512	1,210	780	684	134	2,808

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: Baseline analysis, 1816-2001**  
(dependent variable: dichotomous indicator of conflict)

	(1)	(2)	(3)	(4)	(5)
	Univariate, probit	Geography controls, probit	Baseline, probit	Baseline, logit	IV probit, instrument- ing with 1500 GD
<b>Fst genetic distance, weighted</b>	<b>-7.4543 (12.297)**</b>	<b>-1.3275 (5.837)**</b>	<b>-0.9313 (8.922)**</b>	<b>-0.7389 (-6.224)**</b>	<b>-1.4414 (-6.511)**</b>
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)**
Pseudo-R2	0.047	0.208	0.300	0.309	-

Robust t statistics in parentheses (clustering at the country pair level);

\* significant at 5%; \*\* significant at 1%.

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

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%)

**Table 4: 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	1914-1945	1946-2001	1919-1989
Fst genetic distance, weighted	-0.9313 (8.922)**	-0.8059 (0.526)	-0.7590 (8.668)**	-0.6467 (0.684)	-0.3915 (7.293)**	-0.7591 (7.372)**
Log geodesic distance	-0.0735 (4.487)**	-0.3183 (2.887)**	-0.0608 (4.247)**	-0.5597 (5.745)**	-0.0216 (2.919)**	-0.0717 (4.225)**
1 for contiguity	0.4227 (7.760)**	0.8359 (3.759)**	0.3893 (7.645)**	0.6932 (2.585)**	0.2395 (7.411)**	0.5238 (7.235)**
# observations	517,251	32,292	484,959	50,281	423,790	33,0365
# of pairs	13,175	799	13,175	2,027	13,175	10,397
Pseudo-R2	0.300	0.186	0.319	0.292	0.346	0.321

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

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

**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 5 - Regional analysis**  
(dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline specification	Same continent dummy	Europe, with Europe FST Gen. Dist.	Asia	Africa	America
Fst genetic distance <sup>a</sup>	-0.9313 (8.922)**	-0.9711 (9.438)**	-49.5053 (1.976)*	-2.1541 (1.824)	-0.6891 (1.755)	-0.2466 (0.383)
Log geodesic distance	-0.0735 (4.487)**	-0.0788 (4.694)**	-0.2775 (1.878)	-0.2850 (4.087)**	-0.0199 (1.488)	-0.0478 (0.865)
1 for contiguity	0.4227 (7.760)**	0.4377 (7.875)**	-0.0090 (0.051)	0.6151 (3.885)**	0.5354 (4.252)**	0.5270 (3.911)**
1 of both countries are on the same continent		-0.0358 (2.380)*				
# of observations	517,251	517,251	22,006	28,738	31,017	35,398
# of pairs	13,175	13,175	291	866	848	581
Pseudo-R2	0.300	0.301	0.266	0.430	0.316	0.296

Robust t statistics in parentheses (clustering at the country pair level); \* significant at 5%, \*\* significant at 1%. Probit marginal effects reported in all columns. For dummy variables, marginal effects are for discrete changes from 0 to 1. All marginal effects were multiplied by 100 for readability.

<sup>a</sup> : Weighted genetic distance in all columns except column (4), where FST genetic distance between plurality groups from the European genetic distance matrix is entered instead.

**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 - Alternative measures of historical distance, including GD (dependent variable: dichotomous indicator of conflict; estimator: probit)**

	(1)	(2)	(3)	(4)
	Baseline specification	Add linguistic distance	Add religious distance	Add religious and linguistic distances
Fst genetic distance, weighted	-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, Fearon			-0.0872 (1.777)	-0.0882 (1.518)
1 - % cognate measure of linguistic similarity, weighted				
# of observations	440,170	440,170	440,170	440,170
# of pairs	10,021	10,021	10,021	10,021
Pseudo-R2	0.292	0.292	0.293	0.293

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

All coefficients multiplied by 100 for readability. The table reports marginal effects from probit estimates.

In addition to reported coefficients, all regressions include controls for log absolute difference in longitudes, log absolute difference in latitudes, number of landlocked countries in the pair, number of island countries in the pair, dummy for pair shares at least one sea or ocean, log product of land areas in square km, dummy for 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 7 - Nonlinearities and sample splits**  
(dependent variable: dichotomous indicator of conflict; estimator: probit)

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Contiguity interaction	Proximity interaction	Major power interaction	Spline	Quadratic
Fst genetic distance, weighted	-0.9313 (8.922)**	-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)**	-0.0750 (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 Gen. Dist * contiguity		0.5419 (1.177)				
Fst Gen. Dist * proximity			0.3511 (1.309)			
Fst Gen. Dist * 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, weighted						-2.9895 (1.915)
# observations	517,251	517,251	517,251	517,251	517,251	517,251
# of pairs	13,175	13,175	13,175	13,175	13,175	13,175
Pseudo-R2	0.300	0.301	0.301	0.311	0.300	0.301

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

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

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

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS on 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 positive casualties	Probit on War dummy variable (conflict intensity=5)
<b>Fst genetic distance, weighted</b>	<b>-0.1685 (8.290)**</b>	<b>0.2536 (0.613)</b>	<b>-0.2757 (7.413)**</b>	<b>-0.1153 (8.055)**</b>	<b>-0.9103 (0.611)</b>	<b>-0.0875 (6.001)**</b>
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 <sup>a</sup>	0.042	0.135	0.042	0.013	0.252	0.311

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

<sup>a</sup>: Pseudo R-squared in column 6.

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 9: Post-1950 analysis, controlling for trade variables and absolute income differences (dependent variable: dichotomous indicator of conflict; estimator: probit)**

	(1)	(2)	(3)	(4)	(5)
	Baseline with common sample	Control for trade terms	Additional trade terms	Baseline sample	Add absolute income difference
Fst genetic distance, weighted	-0.5223 (7.181)**	-0.3247 (5.810)**	-0.3098 (5.485)**	-0.3102 (5.155)**	-0.3107 (5.055)**
Log geodesic distance	-0.0154 (2.107)*	0.0047 (0.721)	0.0043 (0.652)	0.0045 (0.656)	0.0045 (0.662)
1 for contiguity	0.2528 (6.638)**	0.2186 (7.349)**	0.2134 (7.326)**	0.2363 (7.314)**	0.2364 (7.340)**
Log bilateral openness, t-4		-0.0414 (4.813)**	-0.0395 (4.546)**	-0.0382 (3.998)**	-0.0382 (4.002)**
Log multilateral openness, t-4		0.0552 (1.993)*	0.0595 (2.191)*	0.0343 (1.124)	0.0341 (1.119)
Log distance * log mult. openness		-0.0093 (2.656)**	-0.0098 (2.854)**	-0.0071 (1.866)	-0.0071 (1.859)
Log distance * log bilateral openness		0.0054 (5.095)**	0.0053 (4.883)**	0.0051 (4.337)**	0.0051 (4.340)**
Dummy for zero trade, t-4		-0.0159 (2.382)*	-0.0159 (2.431)*	-0.0165 (2.297)*	-0.0166 (2.316)*
Free trade area (full set)			-0.0236 (2.859)**	-0.0244 (2.765)**	-0.0244 (2.788)**
# of GATT members			-0.0147 (3.768)**	-0.0175 (4.067)**	-0.0175 (4.062)**
Abs. difference in log per capita income					0.0265 (0.107)
# of observations	226,357	226,357	226,357	214,993	214,993
# of pairs	9,127	9,127	9,127	9,127	9,127
Pseudo-R2	0.336	0.350	0.353	0.350	0.350

Robust t statistics in parentheses (clustering at the country pair level);

\* significant at 5%; \*\* significant at 1%.

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

**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.