No 2018-11 – September

Working Paper

Somatic Distance, Trust and Trade

Jacques Melitz & Farid Toubal

Highlights

- Somatic distance, or differences in physical appearance, proves to be extremely important in the gravity model of bilateral trade in conformity with results in other areas of economics and outside in the social sciences and this is also true independently of survey evidence about bilateral trust.

- Genetic distance, properly measured, proves to be a significant influence on bilateral trade together with somatic distance but its effect is fundamentally different from that of somatic distance. Somatic distance is also a more robust influence on bilateral trade than genetic distance.

- Though trust, based on questionnaire answers, has no effect on bilateral trade, the only reasonable interpretation is that the effect of trust in the study remains buried in the other cultural variables.

- The impact of somatic distance, like that of the rest of the cultural variables in the study, cannot be attributed strictly to trust. Therefore, somatic Somatic distance distance adds one more to the many reasons for distinguishing between bilateral and multilateral trade.

RESEARCH AND EXPERTISE
ON THE WORLD ECONOMY

Abstract

Somatic distance, or differences in physical appearance, proves to be extremely important in the gravity model of bilateral trade in conformity with results in other areas of economics and outside in the social sciences. This is also true independently of survey evidence about bilateral trust. These findings are obtained in a sample of the 15 members of the European Economic Association in 1996. Robustness tests also show that somatic distance, as well as co-ancestry, has a more reliable influence on bilateral trade than the other cultural variables. The article finally discusses the interpretation and breadth of application of these results.

Keywords

Somatic Distance, Cultural Interactions, Co-ancestry, Trust, Language, Bilateral Trade.

JEL

F10, F40, Z10.

Working Paper
CEPII

CEPII (Centre d'Etudes Prospectives et d'Informations Internationales) is a French institute dedicated to producing independent, policy-oriented economic research helpful to understand the international economic environment and challenges in the areas of trade policy, competitiveness, macroeconomics, international finance and growth.

CEPII Working Paper
Contributing to research in international economics

Editorial Director:
Sébastien Jean

Production:
Laure Boivin

No ISSN: 1293-2574

CEPII
20, avenue de Ségur
TSA 10726
75334 Paris Cedex 07
+33 1 53 68 55 00
www.cepii.fr
Press contact: presse@cepii.fr

RESEARCH AND EXPERTISE
ON THE WORLD ECONOMY

Somatic Distance, Trust and Trade

Jacques Melitz & Farid Toubal[1]

1. Introduction

In sociology, homophily is taken as an established fact: individuals tend to associate and bond with similar others, as in the proverb "birds of a feather flock together." Also, "Homophily in race and ethnicity creates the strongest divides in our personal environments (McPherson, Smith-Lovin, and James Cook (2001, p. 415)." The authors of this oft-cited article go on: "Age, religion, education, occupation, and gender [follow] in roughly that order." Indeed, discrimination based on race and ethnicity is readily apparent in economics too, especially in labor studies. (See Bertrand and Mullainathan (2004) and Edo, Jacquemet, and Yannelis (2017) concerning job applications; Lang and Lehmann (2012) and Borowczyk-Martins, Bradley, and Tarasonis (2017) concerning wage earnings and employment.) Yet in studies of bilateral trade between countries, homophily is mostly absent. It emerged only recently in a highly influential article by Guiso, Sapienza and Zingales (2009) (hereafter GSZ). Since they wrote, two works have picked up on the theme: Spring and Grossmann (2014) (in a critical spirit) and Yu, Beugelsdijk, and de Haan (2014). GSZ introduced somatic distance or difference in physical appearance based on an Italian source: a work by Biasutti in four volumes, dating 1954 (first edition), which summarizes and extends a huge literature on racial differences in physical anthropology. However, while making use of Biasutti, GSZ and the two aforementioned studies, subordinate the whole issue of somatic distance by treating it as affecting bilateral trade essentially via trust: that is, as an instrument for trust in a 2SLS interpretation.[2] In this contribution, we shall instead introduce somatic distance as a direct influence on bilateral trade right from the start and show that it belongs there.

In his path-breaking work on *The Economics of Discrimination* dating 1957, Becker modeled discrimination in the labor market as founded on the distaste of employers for interacting with certain groups of workers. His view has undergone subsequent modification, with allowances for the significance of beliefs about other people and some rational elements in labor market discrimination (for example, Phelps (1972) and Arrow (1973)). But as the literature in the preceding paragraph is witness, his basic position stands: racial prejudice as such retains a

[1] Corresponding authors : Jacques Melitz : ENSAE and CEPII, Paris. Email: j.melitz@hw.ac.uk., Farid Toubal: ENS de Paris-Saclay, CREST and CEPII. Email: ftoubal@ens-cachan.fr.
We would like to thank Anca Cristea, Anthony Edo, Thierry Mayer, José de Sousa and Romain Wacziarg for valuable comments.

[2] More precisely, GSZ (2009) admit that somatic distance may not be an appropriate instrument for trust as a possibility and Spring and Grossmann (2016) pursue this possibility to some extent. But whereas, to their credit, the latter advance beyond GSZ in this respect, they still center heavily on GSZ's 2SLS specification with somatic distance as an instrument.

firm position in explanations of racial discrimination. It is clear that such prejudice partly stems from and partly breeds distrust. But there is no precedent for viewing the prejudice and, broadly, somatic distance in general, as impinging on international trade strictly through the filter of trust. No one would argue that sex discrimination is only a matter of trust. There should be little question either that discrimination based on physical appearance is not strictly a matter of trust. In the specific case of international trade, it is standard, since Armington (1969), to allow that national preferences for different trade partners may intervene in explaining bilateral trade. Should somatic distance be one of the reasons for these national preferences, personal affinities could well be the source, independently of trust. The mechanism could take several forms. For example, an increase in the number of foreign markets to which national firms export raises their fixed costs and necessitates choices. In making these choices, "animal spirits" may operate and thus explain why somatic distance enters. As another example, exporters of consumption goods might find that their wares have more appeal to foreigners who resemble them, partly because of similar tastes, but partly also because of a preference for associating with them in commerce.

Once we admit that somatic distance has a direct place in a gravity equation for bilateral trade, the variable emerges as highly significant. It remains so in the presence of other cultural factors, reflecting language, religion, law, co-ancestry, and the history of wars, as well as sample evidence from questionnaires about trust. Indeed, it is more robust than the rest. Trust, based on questionnaire evidence, and the history of wars never matter. Since immigrants are particularly important in studying cultural interactions, we introduce them. This has a seriously damaging effect on two important cultural variables, same legal origin and common religion. Adding a population-weighted measure of physical distance in the presence of immigrants notably reduces the significance of common native language too. Somatic distance is hardly affected throughout. Somatic distance even outperforms co-ancestry though this last variable also holds up well in a new guise.

All these results occur in a European sample close to GSZ's. There are two strong reasons for sticking close to this sample. The more important is that we want to control for trust and we know no reasonable alternative to GSZ's measure. But secondly, even if we were to drop GSZ's trust variable, we could not extend the analysis very far, only to the rest of Europe outside the European Economic Area (EEA) in 1996, since Biasutti's data for somatic distance permits going no further. This would essentially add Eastern Europe and Switzerland.

As indicated, both somatic distance and co-ancestry, or two different aspects of genetic distance, appear important. The two variables matter singly and jointly. Yet our emphasis will be solely on somatic distance, essentially because of its wrongful neglect and its confusion with trust.

The next section offers the test evidence, the following one provides robustness tests concerning the significance of somatic distance, and the last one engages in general discussion and interpretation.

2. Tests and evidence

The theoretical basis for the gravity model of international trade is now sufficiently well known to permit us to pass directly to the estimating equations. As mentioned, our sample size depends on GSZ's trust variable. They drew this variable from a number of Eurobarometer surveys of the trust of people in one country in natives of another in the then-current EEA. The exact question was: "I would like to ask you a question about how much trust you have in people from various countries. For each, please tell me whether you have a lot of trust, some trust, not very much trust, or no trust at all." The surveys took place in 8 separate years from 1970 to 1996 (there have been no further surveys since) and they cover 15 countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, and the United Kingdom). Of the 8 surveys, GSZ retained 7, and only 5 of the 15 countries appear in all 7, 4 of them only once. As a result, they could only draw 595 observations at most. While adhering to their country sample for the same years as they, we draw more, 690, because they include one variable limiting their sample that we shall neglect (Press coverage). GSZ are also careful to filter out country-specific fixed effects in the raw data to focus strictly on *bilateral* trust. They further eliminate variations in trust by calendar year affecting the entire sample. We shall imitate both steps. Admittedly, their index of bilateral trust is not the only possible one. A notable alternative would be the first difference in two countries' trust in all foreigners as a group (from the World Value Surveys). However, this other index does not pertain to one country's trust in a particular other one, as GSZ's does, and therefore is much poorer.

As a start, the estimating equation is:

$$\text{Ln Exports}_{jit} = \alpha + \beta \text{ Ln Distance}_{ij} + \delta \text{ Border}_{ij} + \gamma \text{ Trust(Q)}_{ijt} + \eta_{it} + \lambda_{jt} + u_{ijt} \quad (1)$$

Ln Exports_{jit} is the log of the exports of country *j* to country *i* in survey year *t*. Trust(Q)_{ijt} is the trust of country *i*, the importer, in country *j*, the exporter, in the year *t* based on questionnaire evidence (Q). η_{it} and λ_{jt} are importer-year and exporter-year fixed effects. u_{ijt} is the residual. For the bilateral exports data, we relied on UN COMTRADE. For distance, we followed GSZ in taking the distances between the two capitals. Common border is the usual 0-1 dummy variable. There are no zeros for the dependent variable. Eq. (1) is there to show the impact of Trust, based on the questionnaire evidence, on trade in the absence of any other cultural variables. It is the sole reflection of any cultural influences in the equation. All the relevant descriptive statistics are reported in the Appendix, Table A1.As seen from the test of eq. (1) in column 1 of Table 1, Trust(Q) is totally insignificant. Distance and common border are very significant, as generally true, but the coefficient of Distance is far below the usual value of one or over (in absolute terms) in the gravity model, and the coefficient of common border is unusually high relative to distance.

Next, we repeat the same test for Somatic Distance after substituting this distance for Trust(Q). As regards the measure of somatic distance (drawn from Biasutti 1954, vol.2), let us quote GSZ in full (GSZ 2009, p. 1107):

Table 1. Baseline results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance (log)	-0.4965***	-0.4006***	-0.3956***	-0.4133***	-0.3994***	-0.4073***	-0.3401***	-0.3823***
	(0.0914)	(0.0860)	(0.0775)	(0.0771)	(0.0754)	(0.0772)	(0.0752)	(0.0757)
Common border	0.7246***	0.6341***	0.2977***	0.2775***	0.2897***	0.3288***	0.3024***	0.3181***
	(0.1488)	(0.1256)	(0.1070)	(0.1046)	(0.1055)	(0.1057)	(0.1038)	(0.1049)
Trust(Q)	0.3043		-0.2609	-0.2062	-0.2058	-0.0429	-0.2066	-0.0088
	(0.1969)		(0.1799)	(0.1753)	(0.1753)	(0.1719)	(0.1688)	(0.1683)
Common official language			0.4762***	0.0407	0.0607	-0.0674	0.0270	-0.1869
			(0.1508)	(0.1700)	(0.1811)	(0.1813)	(0.1715)	(0.1668)
Common native language				1.0398**	1.0204**	1.1985**	1.0944**	1.3711***
				(0.4519)	(0.4550)	(0.4718)	(0.4587)	(0.4879)
Same legal origin			0.3079***	0.2447**	0.2309**	0.2231**	0.2115**	0.2377**
			(0.0902)	(0.0950)	(0.0920)	(0.0941)	(0.0959)	(0.0958)
Common religion			0.2616*	0.3469**	0.3618***	0.4049***	0.3886***	0.4032***
			(0.1386)	(0.1438)	(0.1379)	(0.1377)	(0.1403)	(0.1395)
Linguistic common roots			0.3334	0.2850	0.2778	0.4330	0.3801	0.5983**
			(0.2672)	(0.2706)	(0.2757)	(0.2721)	(0.2672)	(0.2500)
Somatic distance(HHC)		-0.1330***	-0.0994***	-0.0905***	-0.0842***		-0.0853***	
		(0.0323)	(0.0305)	(0.0302)	(0.0307)		(0.0300)	
Co-ancestry(1)					0.1654	0.5471		
					(0.4236)	(0.4126)		
Co-ancestry(2)							0.4629**	0.5047***
							(0.1856)	(0.1925)
Observations	690	690	690	690	690	690	690	690
Adj. R2	0.947	0.953	0.963	0.963	0.963	0.962	0.964	0.963
Number of country pairs	207	207	207	207	207	207	207	207

Notes : The dependent variable is the log of aggregate exports from country i to country j. All regressions contain exporter/year and importer/year fixed effects. Standard errors are in parentheses. These are based on robust standard errors that have been adjusted for clustering by country pair. Coefficients are statistically different from zero at the ***1%, **5%, and *10% level.

As an alternative measure of distance [to the DNA measure] between two populations, we derive an index of somatic distance, based on the average frequency of specific traits in the indigenous population reported by Biasutti (1954). For height, hair color (pigmentation), and cephalic index (the ratio of the length and [to the] width of the skull). Biasutti (1954) draws a map of the prevailing traits in each country in Europe. For each trait, European Union countries fall into three different categories. For hair color we have "Blond prevails," "Mix of blond and dark," and "Dark prevails." We arbitrarily assign the score of 1 to the first, 2 to the second, and 3 to the third. When one country's somatic characteristics belong to more than one category, we take the country's most prevalent category. We then compute the somatic distance between two countries as the sum of the absolute value of the difference between each of these traits.

Column 2 shows the result of substituting Somatic Distance for Trust(Q). As we see, Somatic Distance is extremely important with the right sign and the estimates for physical distance and common border are moderately lower than before (in absolute terms), but remain highly significant.

Following, we use a mix of Trust(Q) and Somatic distance, while adding a range of controls for other cultural influences besides the obvious one of common language. As regards these other controls, GSZ make a whole series of interesting suggestions. They introduce five variables, all of them possibly for the first time in the gravity literature: namely, first, same legal origins; second, the history of wars between countries going back to the year 1000; third, common religion; fourth, common linguistic roots based on the *Ethnologue* classification of language trees; and last, a different measure of genetic distance besides somatic distance depending on DNA sequences, which in accordance with the literature, we will label co-ancestry. For Common language, they resort to Common official language. Their measure of Same legal origin comes from La Porta, López-de-Silanes, Schleifer and Vishny (1998), who distinguish between French, German, Scandinavian, and English origins. These first two are dummy variables. The history of wars will not detain us since GSZ dropped the variable early on because it proved insignificant (as has not always been true since) and we do too. Common religion comes from the *World Value Surveys* of the World Bank, which distinguish between Catholic, Protestant, Jewish, Hindu, Buddhist, Orthodox, no religion and other affiliation. Common religion is also a 0-1 indicator variable. Linguistic common roots rest on the Fearon-Laitin (2003) index based on language trees. Co-ancestry rests on Cavalli-Sforza, Menozzi, and Piazza (1996) and measures the degree to which selected CNA sequences ("markers") are identical in the genetic analysis. A rise in co-ancestry reduces genetic distance. We tried alternative measures for all these variables before deciding to stick to GSZ's with two exceptions, concerning language and co-ancestry, which we will justify.

Column 3 shows the results of adding the aforementioned cultural influences except for the two that GSZ drop early on: the history of wars and co-ancestry. They drop DNA sequences early on as well as the history of wars because somatic distance dominates the variable completely when both measures of genetic distance serve together. As seen from column 3, the coefficient of Trust(Q) is still insignificant, as in column 1, but with the wrong negative sign.

Common official language and Same legal origin both enter positively and very significantly at the 99 percent confidence level. Common religion does the same but only at the 90 percent confidence level. Linguistic common roots is totally insignificant. Finally, Somatic distance remains negative and significant at the 99 percent confidence level just as before in column 2.[3]

Column 4 focuses on language. At the time GSZ wrote, the only measure of common language in wide use rested on official status. Rose (2000) had recently pioneered this measure in applying the gravity model to worldwide evidence. Widely serviceable measures of common language based on native language and spoken language only came shortly after. Since they came (or concurrently), Melitz and Toubal (2014) have shown the superiority of both of these measures to official language in measuring a common language. The point bears special note at present. In the current sample of the EEA membership of 1996, for example, English is an official language strictly in the United Kingdom and Ireland and therefore, based on official status, English is a means of communication strictly between these two. As another example, German is official strictly in Germany, Austria, and Switzerland; yet we know of the importance of German in the Netherlands, Denmark and the Scandinavian countries.

In light of these concerns, we simply add Common native language in column 4 to the rest of the variables in column 3 of Table 1. In order to do so, we constructed a special measure of the variable to adapt it to the study period 1970-1996. Since most of the observations in this period come in the latter part, the new measure rests on the information about native languages in the 15 relevant countries in the 1988, 1996 and 2000 editions of *Ethnologue*. First, we averaged the percentages of native speakers of the relevant languages over those three years. Next, we calculated the sum of the *products* of the population shares country pair by country pair. (We needed to sum the products because a country pair could contain native speakers of both German and French in both countries, for example.)[4] The resulting measure refers to the probability that two people at random from two different countries in the sample would have the same native language.

[3] GSZ obtain moderately better results for Trust(Q) in their OLS estimates than ours, but the differences are easy to explain. They stem from two sources: first, GSZ's addition of Press coverage, costing many observations (and perhaps secondarily their addition of Transportation costs); and second, their exclusion of somatic distance and common religion from their OLS estimates of bilateral trade in order to reserve both for use as instruments for trust(Q) in a subsequent 2SLS specification. If we replicate their OLS equations in the first three columns of their Table IV (GSZ 2009, pp. 1116-7), we get a somewhat higher coefficient (0.42 instead of 0.36) and higher standard error (0.21 instead of 0.17) for Trust(Q) than they do in their column 1, a slightly lower coefficient (0.27 instead of 0.29) and higher standard error (0.19 instead of 0.17) in their column 2, and a slightly lower coefficient (0.22 instead of 0.25) and the identical standard error (0.19) in their column 3. As thus clear, our estimates and theirs for Trust(Q) are quite close on their specification. In addition, if we introduce Press coverage (thereby losing nearly 100 observations) and transportation costs in our own specification, which contains both common religion and somatic distance, none of our conclusions is affected.

[4] Admittedly, this procedure can lead to a problem of double-counting because of bilinguals and trilinguals, etc., but that is only an important concern for common spoken language not for common native language (see Melitz and Toubal 2014, p. 354).

As seen from column 4, Common native language enters significantly at the 95 percent confidence level. In its presence, the significance of Common official language vanishes completely. Otherwise, there is little change except that Same legal origin drops in significance to the 95 percent confidence level and Common religion rises to the 95 percent confidence level.

The next four columns, 5 through 8, center on co-ancestry. Column 5 shows that the variable, as measured by GSZ and labelled Co-ancestry(1), is insignificant, just as GSZ say (cf. Giuliano, Spilimbergo and Tonon 2014). Next, in column 6, we allow the variable to stand as the sole reflection of genetic distance by removing Somatic distance. Co-ancestry(1) remains totally insignificant. These results may seem to go contrary to the recent literature stemming from Spolaore and Wacziarg (2009, 2013, 2016) which emphasizes DNA measures of co-ancestry as an important factor in many economic contexts. As the note below explains, this is not clear.[5] However, the whole issue has been recently superseded by the appearance of a new measure of co-ancestry in genetics, superior to the old one, according to which co-ancestry is indeed important.

Cavalli-Sforza et al (1996) provided bilateral genetic distances between 42 populations. Recently, Pemberton, DeGiorgio, and Rosenberg (2013) combined eight datasets appearing since Cavalli-Sforza et al to construct a new measure of bilateral genetic distances covering a grid of 267 worldwide populations. Equally important, this newly compiled dataset reflects DNA sequences at the molecular level (microsatellite variation) as was not true before. These DNA sequences have much higher mutation rates and diversity than the earlier ones. The result is to provide finer distinctions on a much wider world scale. Since Pemberton *et al* wrote Spolaore and Wacziarg (2018) have adapted the former's dataset, pertaining to ethnic pairs, to apply to country pairs. Once Spolaore and Wacziarg apply their adapted measure of co-ancestry to their previous work, they report: "[the] results come out strengthened in terms of standardized magnitudes and levels of significance" (2018, p. 2). We obtain even greater improvement than they. Whereas Co-ancestry(1) had been totally insignificant before in our work, the new measure, Co-ancestry(2), performs well.[6]

When we substitute Co-ancestry(2) for Co-ancestry(1) in column 7, we see that the variable becomes important at the 95 percent confidence level while the rest of the equation, including Somatic distance, is barely affected (except possibly for Common native language, whose coefficient falls notably but whose significance stays about the same). In column 8, we drop

[5] The relevant literature using co-ancestry(1) may not conflict with our results because this literature typically does not consider bilateral trade and when it does, as in Giuliano et al (2014) (who also study a sample of European countries, though larger than ours), with one exception (Bove and Gokmen 2018), co-ancestry(1) is unimportant. Giuliano et al (2014) emphasize the role of geography, including mountains, average elevation and access to seas – apart from distance and common borders – instead of DNA evidence in explaining the impact of ancestry on bilateral trade. In addition, somatic distance never appears as a separate variable in this literature and we do not know what the result would be otherwise.

[6] We drew our co-ancestry(2) dataset for our 15-country sample from Spolaore and Wacziarg's population-weighted measure (as opposed to their unweighted one), both of which were based on Pemberton *et al* (2013). For another use of Spolaore and Wacziarg's recent work, see also Fensore, Legge and Schmid (2018).

Somatic distance and retain Co-ancestry(2) as the only measure of genetic distance. The coefficient of Co-ancestry(2) now rises mildly and attains significance at the 99 percent confidence level (to which it was close before). The biggest change relates to language. Common native language becomes significant at the 99 percent level. Even more strikingly, without Somatic distance, Linguistic common roots is important for the first time. In view of this significance of Co-ancestry(2), we will return to the distinction between this variable and somatic distance, our primary concern.

Column 7 is our preferred estimate in Table 1. According to it, native language is the largest of the five significant cultural influences on trade. Based on standardized beta coefficients, its impact is 8 percent, that of Co-ancestry is 7.5 percent, and those of Same legal origin, Common religion and Somatic distance are 6 percent.[7]

3. Robustness tests

For the rest, our econometric analysis centers on the robustness of Somatic distance, its robustness alone and as compared with the other cultural variables. Table 2 begins with a repeat of column 7 of Table 1 without common official language. In column 2, we take advantage of supplementary information about somatic distance on GSZ's website (GSZ 2017). There, GSZ provide an interesting alternative measure. Whereas the current measure rests on height, hair color, and cephalic index (HHC), the website offers the possibility of testing based on height and hair color (HH) alone. This obviously can shed light on the separate importance of the cephalic index. As we see in column 2, measuring somatic distance based on height and hair color alone makes little difference. We shall continue in the remainder of our tests to show results for Somatic distance resting on both HHC and HH alone.

Next, we introduce a particularly strong robustness test, to our minds: we admit immigrants. Not only does this variable regularly enter highly significantly in previous estimates of bilateral trade, but its presence tends to lower, blur or even eliminates the influence of other cultural variables. Our measure of Immigrants is the stock of people in country *j*, the exporter, who were born in country *i*, the importer (Özden, Parsons, Schiff and Walmsley 2011). This variable (Immigrants) is obviously subject to simultaneity bias since exports from country *j* into country *i* may encourage emigration from *i* to *j*. The reverse influence of trade on Immigrants is thus also expected to be positive. Consequently, failure to correct for it (and we do not know how to do so), in principle, should lead to an exaggerated positive coefficient of the variable. All the other coefficient estimates in the equation could be affected.

[7] We consider Co-ancestry(2) a cultural variable because of ample evidence that it reflects nurture as much, if not more than nature (see Spolaore and Wacziarg 2013, in particular pp. 348-63).

Table 2. Robustness tests

	(1)	(2)	(3)	(4)
Distance (log)	-0.3410***	-0.3520***	-0.3783***	-0.3888***
	(0.0750)	(0.0789)	(0.0669)	(0.0709)
Common border	0.3031***	0.3568***	0.2578***	0.3047***
	(0.1023)	(0.0972)	(0.0923)	(0.0882)
Trust(Q)	-0.2029	-0.2066	-0.0833	-0.0818
	(0.1644)	(0.1780)	(0.1292)	(0.1430)
Common native language	1.1403***	1.3084***	0.7532**	0.8965***
	(0.3173)	(0.3505)	(0.3004)	(0.3348)
Linguistic common roots	0.3808	0.2208	0.1800	0.0465
	(0.2666)	(0.2908)	(0.2443)	(0.2624)
Same legal origin	0.2103**	0.1936**	0.1390	0.1254
	(0.0950)	(0.0967)	(0.0891)	(0.0914)
Common religion	0.3896***	0.3909***	0.2616*	0.2626*
	(0.1402)	(0.1410)	(0.1352)	(0.1347)
Co-ancestry(2)	0.4633**	0.4547**	0.3534**	0.3468**
	(0.1852)	(0.1994)	(0.1600)	(0.1700)
Somatic distance(HHC)	-0.0846***		-0.0757***	
	(0.0290)		(0.0261)	
Somatic distance(HH)		-0.0940**		-0.0817**
		(0.0378)		(0.0340)
Immigration (log)			0.1301***	0.1302***
			(0.0282)	(0.0285)
Observations	690	690	690	690
Adj. R2	0.965	0.964	0.969	0.968
Number of country pairs	207	207	207	207

Notes: The dependent variable is the log of aggregate exports from country i to country j. All regressions contain exporter/year and importer/year fixed effects. Standard errors are in parentheses. These are based on robust standard errors that have been adjusted for clustering by country pair. Coefficients are statistically different from zero at the ***1%, **5%, and *10% level.

With these caveats in mind, column 3 shows what happens when the variable enters. As ever, Immigrants is extremely important. In its presence, Same legal origin ceases to matter at all. The coefficient of Common religion drops from the 99 percent confidence level below the 95 percent one. The coefficient of Common native language drops as well, while the variable remains important at the 95 percent confidence level. Of all the cultural variables, only Somatic distance (HHC) and Co-ancestry(2) are essentially unaffected. Their coefficients drop moderately and their significance remains about the same. The last column of Table 2 substitutes Somatic distance (HH) for Somatic distance (HHC). HH behaves mildly worse than HHC, as before in the first two columns. In addition, with HH present instead of HHC, Common native language improves. Otherwise, there is little change to speak of.

Table 3 takes up an important suggestion in Spring and Grossmann (2016, pp. 107-108). They observe that Head and Mayer (2002) had argued for the use of population-weighted distances in limited regional samples like ours on the ground that the other measures of distance tend to exaggerate the effects of Common border. We find, as Spring and Grossmann did, that use of the population-weighted measure gives more plausible results in the GSZ sample.

The first two columns of Table 3 begin by removing Immigrants in order to distinguish between the effect of the new measure of distance and that of Immigrants. Let us focus first on Distance and Common border, where the basic change lies. As compared to the first two columns of Table 2, the coefficients of Common border in columns (1) and (2) are hardly affected. On the other hand, the coefficient of Distance rises substantially in absolute terms, from −0.34 (or −0.35) to −0.56 (or −0.57). This approach to −1 strikes us as good reason to favor population-weighted distance since −1 or higher (in absolute terms) is more typical in gravity tests. Viewing the cultural variables next, we find that the switch to population-weighted distance does some light harm. Common language, Common religion, Co-ancestry(2) and both Somatic distance(HHC) and (HH) suffer a drop in coefficients. Concomitantly, Common religion and Somatic distance(HHC) drop in significance from the 99 to the 95 percent level, and Co-ancestry(2), drops from the 95 to the 90 percent confidence level when combined with Somatic distance(HH) instead of somatic distance(HHC) (columns 2 of Tables 2 and 3). Only Same legal origin is relatively unaffected.

The picture changes when population-weighted distance and Immigrants are mixed together; now the results are far worse. As we see in the last two columns of Table 3, Common religion becomes insignificant for the very first time. Same legal origin remains significant at the 90 percent confidence level in column 3 but not column 4. Common native language suffers considerably too, dropping in significance from the 99 to the 90 percent level in both columns. Co-ancestry(2) and Somatic distance(HH) also do more poorly. Only Somatic distance(HHC) stands up well.

Table 3. Population-weighted distance

	(1)	(2)	(3)	(4)
Pop-Weighted Distance (log)	-0.5580***	-0.5719***	-0.5517***	-0.5662***
	(0.1049)	(0.1103)	(0.0973)	(0.1036)
Common border	0.2644***	0.3068***	0.2427***	0.2802***
	(0.0965)	(0.0916)	(0.0904)	(0.0877)
Trust(Q)	-0.2554	-0.2537	-0.1487	-0.1423
	(0.1555)	(0.1658)	(0.1275)	(0.1386)
Common native language	0.8643***	0.9872***	0.5270*	0.6320*
	(0.3256)	(0.3650)	(0.3169)	(0.3575)
Linguistic common roots	0.2435	0.1221	0.0949	-0.0065
	(0.2576)	(0.2849)	(0.2442)	(0.2644)
Same legal origin	0.2207**	0.2087**	0.1506*	0.1413
	(0.0924)	(0.0940)	(0.0875)	(0.0897)
Common religion	0.3199**	0.3195**	0.2119	0.2115
	(0.1393)	(0.1397)	(0.1344)	(0.1336)
Co-ancestry (2)	0.3789**	0.3727*	0.3179**	0.3130*
	(0.1800)	(0.1903)	(0.1574)	(0.1656)
Somatic distance(HHC)	-0.0699**		-0.0643**	
	(0.0270)		(0.0249)	
Somatic distance(HH)		-0.0743**		-0.0656*
		(0.0371)		(0.0342)
Immigration (log)			0.1143***	0.1143***
			(0.0277)	(0.0281)
Observations	690	690	690	690
Adj. R2	0.966	0.966	0.969	0.969
Number of country pairs	207	207	207	207

Notes : The dependent variable is the log of aggregate exports from country i to country j. All regressions contain exporter/year and importer/year fixed effects. Standard errors are in parentheses. These are based on robust standard errors that have been adjusted for clustering by country pair. Coefficients are statistically different from zero at the ***1%, **5%, and *10% level.

As we look back on the entire set of results in Tables 1, 2, and 3, Somatic distance behaves distinctly better than the other cultural variables except possibly Co-ancestry(2). Apart from the outlying value of −0.13 when the variable served as the sole reflection of any cultural influences (Table 1, column 2), its coefficient varies from −0.10 to −0.06 from start to finish, and its significance never falls below the 95 percent confidence level to the 90 percent one except once when it is measured strictly on the basis of height and hair color (Table 3, column 4). By contrast, all of the other cultural variables but Co-ancestry(2) truly wilt at some point. Common language does so in the last two columns of Table 3 when it barely remains important. Same legal origin does so in the last two columns of Table 2 and the last column of Table 3 when it becomes unimportant. Common religion moves from excellent performance at first in Table 2 to fair performance in the last two columns of the table, and from good performance in Table 3 to poor performance in the last two columns of the table. Upon examination, Somatic distance(HHC) also outperforms Co-ancestry(2). It never falls below significance at the 99 confidence level in the first two tables whereas Co-ancestry(2) never attains this level in all three tables except once, namely, when Somatic distance drops out entirely in Table 1, column 8. In addition, Somatic distance(HHC)'s low is 95 percent significance whereas that of Co-ancestry is 90 percent. Finally, we are also prone to attach importance to the consistently better performance of HHC than HH in Tables 2 and 3 and thus the presence of the cephalic index in measuring Somatic distance, though HHC's advantage is never large.

4. Discussion

General discussion may begin with the joint significance of the two best-performing measures of genetic distance, Somatic distance(HHC) and Co-ancestry(2). This joint significance implies that the impact of somatic distance is clearly distinguishable from that of co-ancestry, though both variables reflect genetic distance. Sticking to strict statistics, this is not surprising, since the two variables are totally uncorrelated. Their simple correlation is 0.04 (after introducing country-year fixed effects their correlation rises to 0.11).[8] A look at the full correlation matrix in the appendix (Table 1B) shows that there is more reason for concern about the ability to separate the impact of either variable from the other cultural influences than from one another. Note too that when Co-ancestry(2) enters in column 7 of Table 1, the coefficient of somatic distance drops little in absolute terms (from its previous level in column 4) and stays equally significant, and when Somatic distance exits in column 8 of this table, the coefficient of Co-ancestry(2) rises little (absolutely) from its level in column 7 and becomes barely more significant (though nevertheless crossing from the conventional 95 percent level of significance to the 99 percent one).

Yet we cannot leave the matter here. The marked independence of somatic distance and co-ancestry in our results is not exactly intuitive. To delve further, Spolaore and Wacziarg (2009, 2013, 2016) explain that the contemporary effects of co-ancestry come from the survival of

[8] This independence of Somatic distance(HHC) and Co-ancestry(2) may be partly due to the limitation of the sample to Europeans and would diminish in a sample comparing Europeans with Asians, Africans, and Latin Americans. But we do not know. The use of Co-ancestry(2) instead of co-ancestry(1) may also be a factor. GSZ (2009, p. 1108) report a correlation coefficient of –0.53 between Co-ancestry(1) and Somatic distance(HHC) in the presence of country-year fixed effects. We found virtually the same high coefficient in absolute terms, –0.515, given these fixed effects. However, if we remove the fixed effects, the correlation coefficient between Co-ancestry(1) and Somatic distance(HHC) drops in absolute terms to –0.12.

common human traits. The less the time since pre-history when two populations divided from their common ancestors, the less time the two have had to deviate from one another genetically: thus, as measured, the higher their co-ancestry. According to the evidence, a higher transmission of genetic traits from generation to generation has numerous current manifestations, extending from output performance to responses on World Value Surveys to questions about personal values and generalized trust. The usual explanation of these (not necessarily evident) linkages from the very distant past traces many of them to the persistence of aspects of family, social and political organization over centuries and millennia, with reference to and support from, major historical effects along the way, for example, on the timing of adoption of agriculture and the speed of transfer of technologies after the industrial revolution. Yet the measures of co-ancestry themselves relate to things like the sharing of blood types and resistance to different viruses that cannot even be detected without sophisticated instruments. They say little or nothing about phenotypes, including hair color, height, and cephalic dimensions. By contrast, somatic distance presents itself to the eye. Its impact is easily associated with such other current social features as homophily and racial discrimination without necessarily invoking the distant past. This then explains why somatic distance could operate largely through different channels than co-ancestry and affect bilateral trade largely independently, just as our statistical results say is the case.

Next, we may turn to the insignificance of Trust(Q) in explaining bilateral trade. On a strictly formal level, there are two possible interpretations. One is that trust has no bearing on bilateral trade. The other is that Trust(Q) is a poor estimate of trust in bilateral trade and therefore that trust remains embedded in the other cultural influences (as a group) in the estimates. There is little doubt that the second alternative is to be preferred. It would be unreasonable to deny the importance of trust in trade simply because of poor results in a limited sample of answers to questions about trust (subject to sample variance and demanding respondents to put up no stake whatever) in the teeth of all the previous historical evidence, much of it going back many centuries, to the contrary. Nor can we easily overlook the theoretical and common sense grounds for this historical evidence to the contrary (see Algan and Cahuc 2015 for good discussion and summary). Significantly, though, the answer has a drawback. It forces us to recognize that Somatic distance, however independent it may be of common language, common religion, common law, and co-ancestry in our results, most likely reflects an element of trust. This ambiguity is difficult to remove. However, the same ambiguity is also difficult to remove from the impact of common language, common religion, common law, and co-ancestry. Yet except for GSZ regarding common religion,[9] to our knowledge no one pins these other cultural variables' impact exclusively on trust. Is there really any reason to do differently for somatic distance? Related evidence from sociology, labor economics, psychology and game theory,[10] says no. According to this evidence, social preferences, biases and

[9] As mentioned before (note 2), GSZ use common religion as well as somatic distance as instruments for trust. See Fehr (2014, p. 259) for earlier criticism of this use of common religion.

[10] We cite psychology and game theory here alongside sociology and labor economics because of evidence from these other fields of study of the presence of differences in altruism and cooperation depending on in-group status (see Fehr 2014). Interestingly too, even under laboratory conditions of contrived trust games, psychologists and game theorists have

discrimination are very likely to underlie people's responses to others' physical appearance independently of any rational calculations of expected returns.

Next, what shall we make of our better results for somatic distance than all the other cultural variables? How much credence can we give to this outcome? Rather little, we think. In this case, the limitation of our sample to the 15 European members of the EEA in 1996 is a big obstacle. The importance of this limitation shows up in numerous ways. One is the impact of substituting population-weighted distance for distance between national capitals. Much previous work shows that once a sufficiently wide world sample of countries serves, it makes no difference in gravity equations whether distances are measured based on capitals, central geographical locations, most populated cities, or otherwise. If population-weighted distances rarely serve in tests of *world* samples, as they seem to do, it is only because researchers know that the added sophistication would be pointless. We have checked this point in our 193-country sample in Melitz and Toubal (2014) by substituting population-weighted distance for distance between most-populated cities, our earlier choice. The differences in the coefficients and standard errors for distance and common borders and all the rest of the variables are hardly worthy of discussion.

As a second indication of the importance of our limited sampling, past work with worldwide samples also tends to yield better results for common language than those here. In a study focusing on the proper measurement of common language, we found, in sharp contrast to the present, that common official languages, common native languages and differences in linguistic roots all emerge as simultaneously important (ibid.). True, somatic distance was absent in this earlier work, but all the other cultural variables besides trust and co-ancestry here present are there too.

Common religion and Same legal origin deserve a separate word too. Right now, the only possible common religions are Catholic and Protestant. In a world sample, there would be room for other important shared religions such as Muslim, Buddhist and Orthodox. The results could be better or worse. Similarly, GSZ's measure of Same legal origin, which distinguishes between French, German, Scandinavian, and English origins, would be difficult, if not impossible, to apply on a global scale. Any global measure of legal systems would need to rest on a different classification, for example, JuriGlobe's between civil law, common law, Muslim law, and mixed systems. Once again, the variable might behave differently. For all these reasons, we must beware of concluding that Somatic distance is the most reliable of the cultural influences.

Still, what about the importance of somatic distance as such? In our view, the variable's behavior in our tests provides powerful evidence in its favor. One might have thought that in a test controlling for the ability to communicate through language (not simply official language), common law, common religion, common linguistic roots, and co-ancestry, the significance of

great difficulty distinguishing effects of trust from ones of risk aversion and philanthropy (or a mix of personal ethic and self-esteem).

the variable could not appear. But the results show otherwise. In the presence of all of these cultural factors, comfort in dealing with others who look more like ourselves carries over to foreign trade. We gave earlier examples of the possible mechanisms through which this may happen. In our European sample, the influence on trade almost surely means more than simply racial discrimination. But in light of the pervasive significance of homophily in social life, this raises no problem.[11]

Independently, it might have seemed that the significance of somatic distance could not appear in our limited sample of 15 European countries because of insufficient variance. However, the results fly in the face of this other prior. In fact, height and hair color alone do almost as well as the variable with a cephalic index too. Furthermore, with or without a cephalic index, Somatic distance outperforms the other cultural variables. Upon separate examination of the issue of variance as such, the three relevant European maps in Biasutti (vol. 2 of the 3d edition, 1959, insets between pages 40-41, 42-43, and 48-49) permit sorting each of the three elements of Somatic distance into three separate groupings, and based on GSZ's method of scoring, the coefficient of variance of Somatic distance is about 0.5 with HHC and .67 with HH alone. Those figures are high enough, it seems, to answer any puzzlement about our ability to discern the impact of the variable. Of course, with wider geographical sampling, there would be still more groupings of each element (that would even be so for all of Europe alone),[12] which would mean more scope for Somatic distance to bear its influence and greater ease of detecting its impact. Once again, based on pure statistics, this could work in any direction. But empirically speaking, it would be surprising if effects of differences in physical appearance on trade that appear clearly in European evidence alone (and in the presence of our many controls), were to blur or disappear with added evidence from Africa, the Middle East and Asia. On the contrary, we would expect the importance of the variable to show up better with broader sampling.

There remains the vexing issue of the absence of skin color in the measure of somatic distance. Many people, we included, would expect this element of somatic distance to be, in fact, the most important. Yet there is no available index for it. A repair of this problem would be very welcome. Perhaps internationally comparable measures of skin color could rest on differences in melanin.

[11] To elaborate, note, as a beginning, that in the context of the gravity model, the issue of the impact of somatic distance on bilateral trade is one, for example, of the Danes' desired trade with the Greeks *as opposed to the Finns and the Swiss and everybody else*. The issue is one of relative affinity for different groups of people in trade. Therefore, varying degrees of homophily from all sources –from attraction to xenophobia – intervene. Likes enter as well as dislikes. We mentioned social science, but the importance of differences in the physical appearance of the "other" is familiar from travel literature and creative writing about cross-cultural encounters too. Somatic distance adds one more to the many reasons for distinguishing between bilateral and multilateral trade.

[12] Biasutti admits more than three possible groupings of height, hair color, and cephalic index, each separately, for Europe as a whole.

References

Algan, Yann and Pierre Cahuc (2014). "Trust, growth and well-being: New evidence and policy implications," in Philippe Aghion and Steven Durlauf, eds. *Handbook of Economic Growth*, vol. 2A, Amsterdam and San Diego: Elsevier, North Holland: 49-120.

Armington, Paul (1969). "A theory of demand for products distinguished by place of production," *International Monetary Fund Staff Papers*, 16: 159-176.

Arrow, Kenneth (1973). "The theory of discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in labor markets*, Princeton: Princeton University Press: 3-33.

Becker, Gary (1957). *The Economics of Discrimination*, Chicago: University of Chicago Press, first edition.

Bertrand, Marianne and Sendhil Mullainathan (2004). "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American Economic Review*, 94(4): 991-1013.

Biasutti, Renato (1954). *Le Razze e i Popoli Della Terra*, vol. 2, Turin : UTET; 3d. ed. 1959.

Borowczyk-Martins, Daniel, Jake Bradley, and Linas Tarasonis (2017). "Racial discrimination in the U.S. labor market: Employment and wage differentials by skill," *Labour Economics*, 49:106-127.

Bove, Vincent and Gunes Gokmen (2018). "Genetic distance, trade, and the diffusion of development," *Journal of Applied Econometrics*, 33: 617-623.

Cavalli-Forza, Luca, Paolo Menozzi, and Alberto Piazza (1996). *The History and Geography of Human Genes*, Princeton: Princeton University Press.

Edo, Anthony, Nicolas Jacquemet, and Constantine Yannelis (2017). "Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications," *Review of Economics of the Household*: 1-28.

Ethnologue. https://www.ethnologue.com/

Fehr, Ernst (2009). "On the economics and biology of trust," *Journal of the European Economic Association,* 7(2-3): 235-266.

Fearon, James and David Laitin (2003). "Ethnicity, Insurgency, and Civil War," *American Political Science Review,* 97(1): 75-90.

Fensore, Irene, Stefan Legge, and Lukas Schmid (2017). "Human barriers to international trade," *University of St Gallen Discussion Paper* no. 2017-12.

Giuliano, Paola, Antonio Spilimbergo and Giovanni Tonon (2014). "Genetic distance, transportation costs, and trade," *Journal of Economic Geography*, 14: 179-98.

Guiso, Luigi, Paola Sapienza and Luigi Zingales (2009). "Cultural biases in economic exchange?" *Quarterly Journal of Economics*, 74(3): 1095-1132.

Guiso, Luigi, Paola Sapienza and Luigi Zingales (2017). "Online appendix for cultural biases in economic exchange?" http://www.kellogg.northwestern.edu/faculty/sapienza/htm/onlineappendix.pdf

Head, Keith and Thierry Mayer (2002). "Illusory border effects: Distance mismeasurement inflates estimates of home bias in Trade," CEPII *Working Papers* no. 2002-01.

Lang, Kevin and Jee-Yeon Lehmann (2012). "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Literature*, 50(4): 959-1006.

La Porta, Rafael, Florencio López-de-Silanes, Andrei Schleifer and Robert Vishny (1998). "Law and finance," *Journal of Political Economy*, 106(6): 1113-1155.

McPherson, Miller, Lynn Smith-Lovin, and James Cook (2001). "Birds of a feather: Homophily in social networks,"*Annual Review of Sociology*, 27: 415-444

Melitz, Jacques and Farid Toubal (2014). "Native language, spoken language, translation and trade,"*Journal of International Economics*, 93(2): 351-363.

Özden, Çağlar, Christopher R. Parsons, Maurice Schiff and Terrie L. Walmsley (2011). "Where on earth is everybody?" *World Bank Economic* Review, 25(1): 12-56.

Phelps, Edmund (1972). "The statistical theory of racism and sexism," *American Economic Review*, 62(4): 659-61.

Pemberton, Trevor, Michael DeGiorgio, and Noah Rosenberg (2013). "Population structure in a comprehensive genomic data set on human microsatellite variation," *G3-Genes/Genomes/Genetics*, 3: 903-19.

Rose, Andrew (2000). "One market one money: The effect of common currencies on international trade," Economic Policy, 15: 7-33.

Spolaore, Enrico and Romain Wacziarg (2009). "The diffusion of development," Quarterly Journal of Economics, 124(2): 469-529.

Spolaore, Enrico and Romain Wacziarg (2013). "How deep are the roots of economic development?" Journal of Economic Literature, 51(2): 325-69.

Spolaore, Enrico and Romain Wacziarg (2016). “Ancestry, language and culture,” in Victor Ginsburgh and Shlomo Weber, eds., The Palgrave Handbook of Economics and Language, London: Palgrave, chapter 7.

Spolaore, Enrico and Romain Wacziarg (2018). “Ancestry and development: New Evidence,” Tufts University Working Paper, March, to appear in Journal of Applied Econometrics, June, https://doi.org/10.1002/jae.2633.

Spring, Eva, and Volker Grossmann (2016). “Does bilateral trust across countries really affect international trade and factor mobility?” Empirical Economics, 50(1): 103-136.

Yu, Shu, Sjoerd Beugelsdijk, and Jakob de Haan (2015). “Trade, trust and the rule of law,” European Journal of Political Economy, 37: 102-115.

Appendix A

Table 1A. Summary Statistics

Variable	Mean	SD	Min	Max
Bilateral Trade	21.25	1.73	15.45	24.79
Somatic distance(HHC)	2.46	1.22	0.00	5.00
Somatic distance(HH)	1.64	1.11	0.00	4.00
Co-ancestry (1)	0.21	0.20	0.00	1.00
Co-ancestry (2)	0.26	0.28	0.00	1.00
Trust(Q)	2.73	0.29	2.01	3.65
Distance (log)	6.92	0.71	5.16	8.12
Common border	0.18	0.39	0.00	1.00
Common native language	0.03	0.12	0.00	0.85
Common official language	0.06	0.23	0.00	1.00
Linguistic common roots	0.37	0.18	0.00	1.00
Same legal origin	0.35	0.48	0.00	1.00
Common religion	0.30	0.26	0.00	0.87
Immigration (log)	9.14	2.07	3.74	13.38

Table 1B. Correlation Matrix

	Linguistic common roots	Common native language	Same legal origin	Common religion	Somatic distance(HHC)	Somatic distance(HH)	Co-ancestry(1)	Co-ancestry(2)
Linguistic common roots	1							
Common native language	0.57	1						
Same legal origin	0.20	0.31	1					
Common religion	0.21	-0.10	0.18	1				
Somatic distance(HHC)	-0.33	-0.19	-0.42	-0.22	1			
Somatic distance(HH)	-0.37	-0.08	-0.37	-0.35	0.85	1		
Co-ancestry(1)	0.47	0.18	0.06	0.27	-0.12	-0.17	1	
Co-ancestry(2)	0.39	0.16	-0.07	0.16	0.04	-0.05	0.45	1