Languages, Genes, and Cultures

Victor Ginsburgh
ECARES, Université Libre de Bruxelles and
Center for Operations Research and Econometrics, Louvain-la-Neuve

July 2004

Abstract

This paper examines three situations in which distances between languages, genes, and cultures matter. The first is concerned with the determinants, which govern the learning of foreign languages. One of these is the "difficulty" of the foreign language, represented by the distance between the native and the foreign language. The second case deals with the formation and breaking-up of nations. Here, it is suggested that genetic distances between regions with diversified populations (such as the Basque country and the rest of Spain) need to be "compensated" by more generous transfer systems if the nation wants to avoid secession-prone behavior. The last case looks at a very popular cultural event, the Eurovision Song Contest, in which nations are represented by singers who are ranked by an international jury that consists of citizens chosen in each participating country. It is shown that what is often considered as logrolling in voting behavior, is rather generated by voting for culturally and linguistically close neighbors.

Published *The Journal of Cultural Econonomics* 29 (2005), 1-17.

1 Introduction

The title of the paper as well as its contents are largely inspired by the work of geneticist Luigi Luca Cavalli-Sforza and his colleagues (1981, 1994, 1995, 2000), though he is certainly not the first in making the link between genes, peoples, cultures, and languages. Charles Darwin (1859) already was aware of some of the connections. Here is what he writes in his *Origin of Species*:[1]

> If we possessed a perfect pedigree of mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world.

An identical view is taken by most linguists, including Joseph Greenberg and Merritt Ruhlen, who trace the evolution and construct the genealogical tree for current languages, as well as by Geert Hofstede who was, to our knowledge, the first to define cultural distances across nations, relating the results of his findings to the historical roots of peoples.

Some reactions during and after my address at the Chicago Association's conference made me aware of the dangers and politically incorrect flavor of the word "gene." It is uncontroversial that genes are important and that, for some reasons that geneticists do not fully understand, we inherit different forms ("alleles") of genes. But this does not mean that there are good genes, and bad genes, with the exception of those that induce some ugly illnesses. Cavalli-Sforza makes this clear in many of his writings. Let me just quote one of these (Cavalli Sforza, 2000, p. viii) and make it mine:

> Genetics is instrumental in shaping us, but so, too, are the cultural, social and physical environments in which we live. The main genetic differences are between individuals and not between populations, or so-called "races." Differences of genetic origins among the latter are not only small (rapidly becoming even smaller with the recent acceleration of transportation, and both migratory and cultural exchange) but also superficial, attributable mostly to reponses to the different climates in which we live. Moreover, there are serious difficulties in distinguishing between genetic and cultural differences, between nature and nurture.

[1] Quoted by Ruhlen (1994, p. 160). See Jones (2000) for a highly entertaining "update" of Darwin's work.

This paper examines three very different situations in which distances between languages, genetic legacies, and cultures do matter in decisions made by rational agents, and shows that these cannot be studied in isolation. The first case is concerned with the determinants that are likely to govern the learning of foreign languages. One of the determinants is the "difficulty" of the foreign language, which is represented by the distance between the native and the foreign language. The second case deals with the formation and breaking-up of nations. Here, it is suggested that larger genetic distances between regions that host populations from very different origins (such as the Basque and Spaniards or Sardinians and Italians) need to be "compensated" by more generous transfer systems if the nation wants to avoid secession-prone behavior. The last case looks at a very popular cultural event, the Eurovision Song Contest, held every year since 1956, in which nations are represented by singers who are ranked by an international jury that consists of citizens chosen in each participating country. It is shown that what is often considered as logrolling in voting behavior, seems rather to be the consequence of voting for culturally and linguistically close neighbors. In the first and third cases, the conclusion results from observed behavior of agents, captured by econometric estimation. In the second case, the approach is theoreticallly grounded, but gives some empirical support to models that are usually rather abstract.

Genetic, linguistic and cultural aspects of a society or a nation are often correlated, since all three are closely linked to nature, but also to learning and history, that is, nurture. One can therefore wonder when it is appropriate to chose one or the other in representing proximity. In some cases, one of the measures is obvious. This is so when describing the difficulty in acquiring a foreign language, though even here, there may be other types of proximities at work between a Swedish and a Danish speaking individual than the mere proximity between the two languages. There is no obvious answer, and certainly no theory to invoke, and the three cases that will be discussed leave this as an open question. Linguistic distances are used in Section 2 to account for the cost of learning a new language. Genetic distances account for the variety of populations that form a nation; this is the topic of Section 3 where the trade-off between secession and transfers is discussed. Linguistic and cultural distances are at play to understand the voting behavior of judges in selecting the winners and the losers in the Eurovision Song Contest considered in Section 4. In each section, the discussion of the notion of distance that is used is followed by the appropriate case study. Section 5 draws some conclusions.

2 The Learning of Foreign Languages

In this section, based on Ginsburg, Ortuno-Ortin and Weber (2004b), we are interested in the determinants that induce people to learn one or several foreign languages. Intuitively, the attactiveness of a foreign language j for a population with native language i depends on the sizes of the two populations. The larger the j-language population (and the smaller the i-language population), the larger the incentive to learn language j for i-natives. But the difficulty to learn the foreign language will also play a role. We represent this difficulty by what linguists define as the distance between languages.

2.1 Linguistic Distances

Pionneered by Morris Swadesh (1952), the lexicostatistical method starts with a list of meanings that are basic enough for every culture to have words for them, for example, *mother, father, blood,* digits, etc. The list used by Dyen et al. (1992) contains 200 such meanings. He collected phonetic representations for the words with these meanings for a group of 94 Indo-European languages.[2] For each meaning, a linguist makes expert judgments of "cognation." Two forms are cognate if they both descend in unbroken lines from a common ancestral word. Words which are common because they have been borrowed are not taken into account.[3] For each pair of languages, a lexicostatistical percentage (or distance) between every pair of languages l

[2] The approach has since be extended to American-Indian, African and other groups of languages.

[3] For example, "flower" was borrowed from the French word "fleur," but "blossom" and "fleur" are cognate. See Dyen et al. (1992, p. 95). Note that ignoring borrowed words often overestimates the distance between languages, since in many cases, such as French and English, a large number of French words have been borrowed by English in the past, while, in more recent times, the borrowing works the other way round. This makes some languages, for instance French and English, less distant than what is suggested by Dyen, though it does not necessarily make their learning easy. Claude Hagège, the French linguist who is Professor at Collège de France, claims that, for a Frenchman, English is "one of the most difficult languages." See *Le Figaro*, April 5, 2004, p. 10.

and m is computed. It is equal to $y_{l,m} = n^0_{lm}/(n^0_{lm} + n_{lm})$, where n_{lm} is the number of meanings for which the speach varieties l and m are classified as "cognate" and n^0_{lm} is the number of meanings for which l and m are "not cognate." The number of "doubtfully cognate" meanings does not enter into the calculation. These percentages are the elements of a matrix of linguistic distances: the larger the number, the more "distant" the two languages. The diagonal elements $y(l, l)$ are set to zero. Table 1 reproduces an example of a distance matrix for some European languages that will be used in the next subsection. Note that distances are symmetric, a strong assumption in this case. It may indeed be easier for a Frenchman to learn English, than for an American to learn French.[4]

It is worth mentioning that this approach has served as basis to regroup languages into family trees which are used to calculate the dates at which separations between languages have occurred. The technique, known as "glottochronology," has, however, recently been seriously challenged by Russel Gray and Quentin Atkinson (2003) whose approach uses computational methods derived from evolutionary biology.

2.2 Modelling the Learning of a Foreign Language

Consider two languages i and j, spoken in two regions or countries i and j, by N_i and N_j citizens. Denote by N_{ij} the number of citizens of country i who study language j. The language utility of an individual depends on the number of those who speak the same language as she does. It is represented by the utility function $U(x, y)$, where x is the (log of) the number of individuals who speak the same native language, while y is the (log of) the number of individuals who share with her a language that is not their native language. Let $n = \log N$. The utility of an i-speaker who learns j is $U(n_i, n_j)$, since she will be able to communicate with all j-speakers. The utility of an i-speaker who does not learn language j is $U(n_i, n_{ji})$: she will communicate with those who know her language in country j. For j-speakers these utilities are respectively $U(n_j, n_i)$ and $U(n_j, n_{ij})$. An individual who learns another language incurs a cost $C(d_{ij})$, where d_{ij} is the (log of) the linguistic distance between

[4]The discussion is largely based on Dyen et al. (1992). For details, see Kessler (2001) as well as Ruhlen (1994), who gives an intuitive approach to the construction of linguistic trees.

languages i and j. Obviously, the larger the distance, the more difficult it is for her to learn the other language.

In a linguistic equilibrium, an individual will be indifferent between learning the foreign language and incuring the cost of learning it, and not learning the language. An (interior) linguistic equilibrium is therefore a solution of the following system of two equations:

$$U(n_i, n_j) - C(d_{ij}) = U(n_i, n_{ji}),$$

$$U(n_j, n_i) - C(d_{ji}) = U(n_j, n_{ij}).$$

We assume that such an equilibrium exists, and leads to demand functions $N_{ij}(n_i, n_j, d_{ij})$ and $N_{ji}(n_j, n_i, d_{ij})$ of individuals whose native language is i (resp. j) to learn the foreign language j (resp. i). Denote by $\log N_{ij}/N_i = F_i(n_i, n_j, d_{ij})$, the (log of the) equilibrium share of individuals whose native language is i and who learn language j. Under some mild assumptions, the function $F_i(n_i, n_j, d_{ij})$ is (a) decreasing in n_i, (b) increasing in n_j and (c) decreasing in d_{ij}.

In order to estimate such demand functions for four of the most important languages (English, French, German and Spanish) of the European Union (EU), we use data that consist of knowledge of native and foreign languages in 13 EU countries,[5] and the distances between languages described earlier.[6]

We estimate a demand function using the following logarithmic functional form for j = English, French, German, Spanish by citizens living in the 13 EU countries $i \neq j$:

$$\log(N_{ij}/N_i)_{EU} = \sum_{j=G,F,S} \alpha_{0j}\delta_j n_j + \alpha_1 n_i + \alpha_2 n_j + \alpha_3 d_{ij} + u_{ij}.$$

The dummies δ_j, j = F, G, and S take the value 1 for observations relative to French, German and Spanish, and 0 otherwise. Each intercept α_{0j} is multiplied by the world population that practices language j. This normalization

[5]Austria, Denmark, Finland, France, Germany, Greece, Italy, Ireland, Netherlands, Portugal, Spain, Sweden, and the United Kingdom. We ignored Belgium and Luxemburg, since in both countries, several languages are "official," which complicates the issue.

[6]See Ginsburgh and Weber (2004) and Ginsburgh, Ortuno-Ortin and Weber (2004a) for details on language proficiency in the EU.

(which does of course not change the other coefficients) makes it easy to test the null hypothesis that all four languages are equally attractive. Note further that the variable $(N_{ij}/N_i)_{EU}$ represents the proportion of inhabitants of EU country i who are proficient in language j, while n_i and n_j represent respectively (the log of) the world populations that are proficient in languages i and j. Estimation results are reproduced in Table 2. They show that (a) the size of the population which learns the foreign language has a negative, though insignificant, effect; (b) the foreign language effect is positive and significantly different from zero; (c) distance between languages carries a negative sign and the effect is significantly different from zero, and (d) the order in which the four languages attract is as follows: English, French, German and Spanish. Though Spanish is widely spoken in Latin America, it attracts less than French and German, that are less widely spoken. This is probably due to the fact that economic reasons and international relations that are not fully captured by the number of speakers also explain why people (have to) learn a foreign language.

It is important to point out that the elasticity of demand with respect to distance is not significantly different from 1, which implies that an increase of 10 percent in the distance decreases with 10 percent the share of natives who learn a given foreign language.

3 The Formation of Nations

There is a growing literature on the formation and breaking-up of nations. Most models try to understand the outcome of two forces that go in opposite directions. The larger the population, the more the common cost of public goods can be shared, and the more transfers between rich and poor regions are possible. This should work as an incentive for diversified nations to stay together. But the larger the population, the more it tends to be linguistically, culturally and genetically diversified. This creates, on the contrary, centrifugal tendencies. In other words, diversity runs against stability but a larger population creates returns to scale which make it possible to decrease the per capita cost of keeping its citizens united.

Most of this literature has remained very theoretical, since "diversification" is often represented by a distribution of the population on a unit segment, *à la* Hotelling-Downs, that is, in a unidimensional world. In their recent paper, Le Breton, Ortuno-Ortin and Weber (2003) suggest to use genetic distances to approximate diversification, which embeds the model into a multidimensional framework or, more precisely, a model in which genetic distances take into account several dimensions.

3.1 Genetic Distances

Populations can be distinguished by their genes, that are strictly hereditary.[7] There exist for instance four blood types (O, A, B, and AB) which are due to large variations of the forms ("alleles") A, B and O of the gene. The frequencies are 27, 8 and 65 percent for Europeans, and 1.7, 0.3 and 98 percent for American natives. The information on frequencies in a single gene is however not sufficient to draw firm conclusions on genetic differences across populations, since such (often different) variations are also present in many other genes. The computation of genetic distances is therefore based on the study of alleles in over one hundred genes in thousands of people from different populations.[8] Table 3 illustrates a few genetic distances, some of which (Basque-Spanish, Sardinian-Italian, and Scottish-English) will be used in the next subsection. Lapps are added to illustrate their remoteness from most other European populations.

Genetic distances based on living populations are used to construct trees to represent the successive separations and migrations among populations over time, and make it possible to go as far as dating these separations.[9] This is thus an idea which is very close to the path taken by glottochronology, and indeed the two sciences (often associated with archeological and paleoanthropological discoveries) have many links, as shown, among others, in the chapter on genes and languages, in Cavalli-Sforza's (2000) book.[10]

[7]The discussion is based on Cavalli-Sforza (2000, pp. 13-25). A much more technical approach is given in Cavalli-Sforza et al. (1994).

[8]There are several formulae for the computation of such distances, but they give reasonably similar results. On this issue, see Cavalli-Sforza et al. (1994, pp. 25-30).

[9]This is due to the fact that genes and allele frequencies change over time according to very precise rules, as do linguistic and phonetic transformations.

[10]See also Cavalli-Sforza et al. (1988).

3.2 The Free Right of Seceding

Le Breton, Ortuno-Ortin and Weber (2003) examine several applications involving the optimal size and composition of a nation. They suggest that the stability of a nation is determined by the diversity of its population, and the cost of governing it.

One of their applications, the one we describe here, considers secession in a country C composed of two regions $i = A, B$. Here is how their model works.

Assume that all citizens contribute identically to the provision of a public good. In general, this requires a transfer from the richer to the poorer region, determined by a redistribution parameter $\beta \in [0, 1]$. The transfer T_i from the country C to region $i = A, B$ is equal to:

$$T_i = \frac{G(P_C)}{P_C} + \beta(Y_i - Y_C),$$

where $G(P_C)$ is the cost of providing the public good to all citizens of C, P_i and Y_i represent the population and GDP per capita in regions i and in country C as a whole.

A citizen of region i incurs two types of costs. The first, due to the diversity of the populations, is equal to

$$\sum_{j=A}^{B} P_j d_{ij}/P_C,$$

where d_{ij} is the genetic distance between the two populations. The second, a consequence of the transfers between regions, is

$$\frac{(y_i - T_i)^{(1-\kappa)}}{1-\kappa},$$

where κ, the absolute degree of risk aversion, is a positive parameter. The disutility of a citizen of region i is defined as a weighted average of the two costs:

$$V_i(C, \beta) = \delta \sum_{j=A}^{B} P_j d_{ij}/P_C + \frac{(y_i - T_i)^{(1-\kappa)}}{1-\kappa},$$

where the weight δ is a given parameter of the society.

The country C that implements the redistribution parameter β is said to be stable under *free right of secession* if for either region $i = A, B$

$$V_i(C, \beta) \leq V_i(i, 0),$$

that is, if the disutility of citizens in either region is larger by seceding than by remaining united.

The form of the function $G(P_C)$ that relates the total cost of the public good to population $G(P_C)$ is set to $\gamma(P_C)^{-1/2}$, which implies that the cost is increasing, but that per capita, it decreases with the size of the population (γ is a parameter).

The goal of the exercise is to find for given parameters γ, δ and κ, the value of the redistribution parameter β that yields stability under free right of secession in three particular cases: the Basque country and Spain, Sardinia and Italy, and Scotland and England. For all three cases, γ is set to 40,[11] while κ is set to 1/2. It is then possible, by numerical calculation, to determine for each δ the value of β that makes the each region indifferent between seceding and staying put.

For a region that is wealthier than the average, there will exist a threshold $\overline{\beta}$ such that for all values below the threshold, there will be no secession. That is, a rich region may tolerate some maximal level of transfers, but its willingness to secede will increase with β. If a region is poorer than the country average, there will also exist a threshold such that for all values of β that are larger than $\overline{\beta}$, there will be no secession. That is, a poor region needs transfers and its willingness to secede decreases with β. Therefore, the secession threshold $\beta(\delta)$ is negatively sloped for regions that are above the country average wealth, and positively sloped for regions below this average.

The Basque country is richer than Spain, so that the slope of the $\beta(\delta)$-locus that makes the Basques indifferent between seceding and not is negative. Since the Basque country enjoys fiscal autonomy, β is roughly equal to 0. Numerical calculations show that there will be no secession as long as $\delta \leq 0.021$.

Calculations lead to a positively sloped indifference locus for Sardinia, since Sardinia is poorer than Italy. For values of $\beta \in [0, 0.4]$ which look

[11] This calibration implies that for a country with a population of 100 and a GDP of 100, the per capita cost of the public good is 4 per cent.

plausible, the implied values of δ vary between 0.020 and 0.015. For values that are smaller, there will be no secession.

Finally, because Scotland and England enjoy very similar GDPs per capita, and the genetic distance between the two populations is very small (27×10^{-4}),[12] the slope of the indifference locus is almost vertical. Stability will occur for any value of the redistribution paramater β as long as δ is smaller than 0.046.

The calibrations show that in all three cases, the utility weight δ put on the diversity component of the utility function has to be rather low for stability to be the rule, especially for the Basque region and for Sardinia. Therefore it comes as no surprise that these regions are much more sensitive to the idea of seceding than is Scotland.

4 Voting Behavior in the Eurovision Song Contest

The Eurovision Song Contest (ESC) was born in 1955, and was held for the first time in Lugano, Switzerland, in 1956, with seven countries competing. At the time, each participant could submit two songs, without any restriction other than the time limit of 3.5 minutes per song. The jury consisted of two delegates from each country who could rate each contestant from 1 to 10. The rules changed several times. Since 1957, every country can chose one song only. In 1958 it was decided that the winning country would host the next contest. The rule concerning the language in which the songs could be performed changed several times, going back and forth between own language only to any language. The number of participants increased from seven in 1956 to 16 in 1961. In 1980, Morocco was added. Egypt came in as participant in 1981, Australia in 1983. In 1996, the number of participants in the finals was limited to 23, then to 24 in 2002. Each ESC is broadcast by television, and since 1985, this happens via satellite. In 2001, the contest was held in Copenhagen in front of an audience of 38,000, and broadcast live all around the world. Nowadays, it is watched by several hundred millions of people.

[12] I am sure that such a small number will be disputed by Sir Alan.

The scoring system changed in 1962, 1964, 1971, and 1975. Since 1975–the first year in our dataset–, the jury members (often a popular jury, not experts), can award rates from 1 to 8, 10 or 12, each number being used only once. This allows 10 songs to be given positive rates. Participating countries cannot vote for their nationals. Until 1997, each country was represented by judges. Televoting was introduced in 1998, so that every citizen can participate, and according to Haan, Dijkstra and Dijkstra (2003), "in many countries, the number of people calling in to register their vote is in the hundreds of thousands." Results in each country are aggregated and announced after all votes have been cast. Participants are ranked according to their aggregate score.

It is frequently said that the vote is political and that countries vote for their neighbors. North European countries (including Scandinavia and the Baltic countries), or South-East European countries (Austria, Romania, Macedonia, Croatia) seem to cast their votes for other members of the group.[13] Some logrolling (that is, exchanges of votes across countries) is also suspected. Ginsburgh and Noury (2004) show that once linguistic and cultural distances are taken into account, the influence of logrolling vanishes. At worst, it seems impossible to distinguish between political and cultural voting. At best, voting is cultural and *not* political.

4.1 Cultural Distances

National culture differences are represented by the five dimensions studied by Geert Hofstede (1980, 1991) and decribed on his website.[14] Hofstede claims that his ideas started with a research project across subsidiaries of the multinational corporation IBM in 64 countries. Subsequent studies by others covered "students in 23 countries, elites in 19 countries, commercial airline pilots in 23 countries, up-market consumers in 15 countries, and civil service managers in 14 countries." These studies identified and scored the

[13]See for instance the discussion on http://homepage.ntlworl.com/waterloo/ 2000/politics.htm, which seems to criticize Terry Wogan, the BBC-TV commentator for having suggested that the vote was political rather than artistic.

[14]I became aware of Hofstede's distances thanks to the paper by McFadyen, Hoskins and Finn (2004) who use them to explain exports of U.S. television programs.

five following dimensions that make cultures differ:[15]

(a) *power distance* measures the extent to which the less powerful members of a society accept that power is distributed unequally; it focuses on the degree of equality between individuals;
(b) *individualism* measures the degree to which individuals in a society are integrated into groups; it focuses on the degree a society reinforces individual or collective achievement and interpersonal relationships;
(c) *masculinity* refers to the distribution of roles between genders; it focuses on the degree to which a society reinforces the traditional masculine work role of male achievement, control, and power;
(d) *uncertainty avoidance* deals with a society's tolerance for uncertainty or ambiguity, and refers to man's search for truth;
(e) *long-term orientation* focuses on the degree to which a society embraces long-term devotion to traditional, forward-thinking, values; long-term orientation is associated with thrift and perseverance, short-term orientation is associated with respect for tradition, the fulfilling of social obligations, and protecting one's face.

The last dimension originates from a research conducted in 23 countries only, and will therefore not be used in our application. This dimension is said to be present in the teachings of Confucius, and is often related to as *Confucian dynamism*.[16]

Hofstede goes explaining that "the grouping of country scores points to some of the roots of cultural differences." Latin countries, for instance, score high on power distance and uncertainty avoidance; this is supposed to come from the time of the Roman empire in which authority was highly centralized, and a law system was applicable to all citizens. The Germanic part of Europe, on the contrary, never suceeded in establishing a lasting common central authority, and countries that inherited its civilization have smaller power distance.[17]

Researchers often use some formulation in which the various distances

[15] The definitions are taken from http://spitswww.uvt.nl/web/iric/hofstede/page3.htm and http://geert-hofstede.international-business-center.com/index.shtml (April 2004), a webpage on which the data can also be found.

[16] See Hofstede and Bond (1988) and Chinese Cultural Connection (1987).

[17] For details, see e.g. Hofstede (1980, 1991).

are aggregated.[18] We use these individually in our calculations, as they seem to represent different components of culture. Since these distances are easily available,[19] we merely give, in Table 4, the example of cultural distances between the United States and some other (mainly European) countries.[20]

4.2 Eurovision: Empirical Results

The purpose is to explain the vote cast by the judges of country $i \in K$ in evaluating the performer of country $j \in K$ ($i \neq j$, since country i cannot vote for its own candidate), where K is the total number of participating countries.

Five groups of factors can be thought of determining the rating:

(a) the quality or intrinsic talent of performers,
(b) the possibility of logrolling on the part of judges,
(c) the order in which artists perform,
(d) other observable characteristics of the performance, and
(e) cultural and linguistic distances between voters and performers.

Talent or quality. Since talent is not observed, the variable must be constructed. One way of doing this is to take the ex-post average rating of a musician j by the juges of country $k \in K$, $k \neq j$. To avoid regressing the vote of judge i on an average which would also include v_{ij}, quality q is defined for each vote v_{ij} cast by i for j as:

$$q_j^i = \sum_{k \neq i,j} v_{kj}.$$

However, an instrumental variables approach sounds more appropriate. Ginsburgh and Noury (2004) thus also estimate a system of two simultaneous equations: the equation of interest, that is, the voting equation, and a second equation in which the quality variable (computed as above) is a function of some exogenous variables.

[18]See, for instance, McFadyen, Hoskins and Finn (2004).
[19]See http://geert-hofstede.international-business-center.com/index.shtml.
[20]The number 70 which measures the distance on masculinity between Switzerland and the United States is dedicated to my Swiss friend Bruno F.

Logrolling or political voting is represented by the vote of the previous year. Thus, the vote of i for j in year t depends on the vote that i received from j in the previous competition.

Order of appearance. The influence of the order in which musicians appear in a competition has often been outlined. Ginsburgh and Flores (1996), Glejser and Heyndels (2001), and Ginsburgh and Van Ours (2003) point this out for one of the top-ranked international piano competitions. Similar observation are made by Haan, Dijkstra and Dijkstra (2003) for the contest that we are dealing with. The exogenous order in which candidates perform is thus also included.

Other observable variables consist of the host country, determined by the citizenship of the previous year's winner–the variable takes the value 1 for the performer whose citizenship is the same as that of the host country–, the language in which the artist sings (English, French, other; recall that the rules changed over time), gender of the artist, and whether the artist sings alone, in a duet or in a group. These are all dummy variables.

Linguistic and cultural distances. The last group of variables includes the linguistic and cultural distances discussed in Section 2.1 and 4.1 (with the exception of long-term evaluation). Table 5 gives the correlations between the various variables for the countries and native languages that are present in our sample. Uncertainty Avoidance is correlated with three other variables, but otherwise, distances seem to pick very different dimensions of peoples' behavior.

The basic equation that is estimated reads

$$v_{ij} = \alpha_0 + \alpha_1 q_j^i + \alpha_2 v_{ji,0} + \alpha_3 O_j + \sum_k \beta_k x_{jk} + \sum_k \gamma_k d_{ij,k} + u_{ij},$$

where v_{ij} is the rating given by country i (the judge) to country j (the contestant), q_j^i is the quality of country j's singer, $v_{ji,0}$ is the rating given by country j to country i in the preceding contest, O_j is the order of appearance of contestant j, the x_{jk} represent "other variables" that characterize j,

$d_{ij,k}, k = 1, ..., 5$ represents the $k-$type distance between countries i and j. The α, β and γ are parameters, and u_{ij} is an error term.

In the simultaneous equations approach, the following system is estimated:

$$v_{ij} = \alpha_0 + \alpha_1 q_j^i + \alpha_2 v_{ji,0} + \alpha_3 O_j + \sum_k \gamma_k d_{ij,k} + u_{ij},$$

$$q_j^i = \sum_k \zeta_k x_{jk} + w_j.$$

Estimation results of the single equation model are described in colunms (1) to (3) of Table 6. Column (4) contains the results concerning the main equation in the simultaneous system, in which quality is instrumented. The first three equations are estimated by Ordinary Least Squares, the fourth by Two Stage Least Squares.[21] Results obtained by Ordered Probit and Tobit estimates are qualitatively identical, and therefore not reported.

We first observe that quality always plays a very significant role, which is of course not surprising.[22] Logrolling is significant (but the value of the coefficient is small) only in Eq. (1), in which no account is taken of linguistic and cultural distances. It remains significant at the 10 percent probability level in Eq. (2) where linguistic distances only are included, but ceases to be so in Eqs. (3) and (4) once other cultural distances are also accounted for. Order of appearance plays no role, while among the other variables of group (d), the only one which has some influence is "sung in French." With the exception of one case, all distance coefficients pick negative signs (the larger the distance, the lower the rating). Linguistic, Uncertainty Avoidance and, to some extent, Power Distance coefficients are very significantly different from zero.

In this popular competition, and contrary to what has often been evoked, there is no evidence for political voting or logrolling, but cultural and linguistic proximities obviously play an important role.

[21]OLS is even possible for the two-equations case, since the system is recursive, though standard deviations would not be the right ones. Note that if the residuals of the two equations are correlated, full information estimation would be more efficient.

[22]Actually, this can be interpreted as signalling that there is large agreement between judges on the rating of candidates.

5 Conclusions

The three cases that were described come from very different backgrounds. All three point to connections between economic behavior (broadly defined), nature, and nurture. People's actions and policies may obviously have an effect on nurture, and in particular, on culture, in the perhaps too broad sense in which it has been extended in this paper. This is, after all, one of the main issues that is dealt with in cultural economics. The three cases analyzed also show that cultural inheritance affects our actions. This is not new either. Max Weber had already pointed out that there may be a thread leading from Protestantism to capitalism, though this is a very disputed idea. Schumpeter dealt with similar issues. So did Georg Simmel in his *Philosophy of Money*–and there is a growing literature on the influence of genes on every component of our behavior. But in many cases the analysis remained essentially sociological and philosophical.[23] "Hard" data on cultural differences, such as distances, have not been used often in cultural economics.[24] Such data are available, and should be more widely lokked at in our field, though, as is argued by Rask (2004), the internet may be knocking distances down.

6 Acknowledgements

I am grateful to Stuart McFadyen and to Shlomo Weber who introduced me to liguistic, genetic, and cultural, distances, and to my coauthors Abdul Noury, Ignacio Ortuno-Ortin and Shlomo Weber who allowed me to borrow from them some joint or non-joint work. I should also like to thank Abdul Noury and Shlomo Weber for comments (on their own work).

7 References

Axelrod, Robert, Michael Cohen and Rick Riolo (2004), Must there be human genes specific to prosocial behavior, *Journal of Economic Behavior and Organization* 53, 49-51.

[23]See however Inglehart and Klingemann (2000), or Axelrod, Cohen and Riolo (2004) for more formal approaches.

[24]An exception is McFadyen, Stuart, Colin Hoskins and Adam Finn (2004), but there may be many other contributions that I am not aware of.

Cavalli-Sforza, Luigi Luca (2000), *Genes, Peoples, and Languages*, Berkeley: The University of California Press.

Cavalli-Sforza, Luigi Luca and Francesco Cavalli-Sforza (1995), *The Great Human Diasporas*, Cambridge, MA: Perseus Books.

Cavalli-Sforza, Luigi Luca and M. W. Feldman (1981), *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton, N.J.: Princeton University Press.

Cavalli-Sforza, Luigi Luca, Paolo Menozzi, and Alberto Piazza (1994), *The History and Geography of Human Genes*, Princeton N.J.: Princeton University Press.

Cavalli-Sforza, Luigi Luca, Alberto Piazza, Paolo Menozzi, and Joanna Mountain (1988), Reconstruction of human evolution: Bringing together genetic, archeological, and linguistic data, *Proceedings of the National Academy of Sciences of the USA* 85, 6002-6006.

Chinese Cultural Connection (1987), Chinese values and the search for culture-free dimensions of culture, *Journal of Cross-Cultural Psychology* 18, 143-164.

Darwin, Charles (1859), *The Origin of Species*, London: John Murray.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black (1992), An Indo-European classification: A lexicostatistical experiment, *Transactions of the American Philosophical Society* 82(5), Philadelphia: American Philosophical Society.

Flôres Jr., Renato and Victor Ginsburgh (1996), The Queen Elizabeth musical competition: How fair is the final ranking? *Journal of the Royal Statistical Society, Series D, The Statistician* 45, 97-104.

Ginsburgh, Victor and Abdul Noury (2004), Cultural voting: The Eurovision Song Contest, manuscript.

Ginsburgh, Victor, Ignacio Ortuno-Ortin and Shlomo Weber (2004a), Disenfranchisement in linguistically diverse societies. The case of the European Union, manuscript.

Ginsburgh, Victor, Ignacio Ortuno-Ortin and Shlomo Weber (2004b), Why do people learn foreign languages?, manuscript.

Ginsburgh, Victor and Jan Van Ours (2003), Expert opinion and compensation: Evidence from a musical competition, *American Economic Review* 93, 289-298.

Ginsburgh, Victor and Shlomo Weber (2004), Language disenfranchisement in the European Union, manuscript.

Glejser, Herbert and Bruno Heyndels (2001), The ranking of finalists in the Queen Elisabeth International music competition, *Journal of Cultural Economics* 25, 109-129.

Gray, Russell and Quentin Atkinson (2003), Language-tree divergence times support the Anatolian theory of Indo-European origin, *Nature* 426, 435-439.

Haan, Marco, Gerhard Dijkstra and Peter Dijkstra (2003), Expert judgment versus public opinion. Evidence from the Eurovision Song Contest, manuscript.

Hofstede, Geert (1980), *Culture's Consequences*, Beverly Hills, California: Sage.

Hofstede, Geert (1991), *Culture and Organizations*, London: McGraw-Hill.

Hofstede, Geert and Michael H. Bond (1988), The Confucius connection: From cultural roots to economic growth, *Organizational Dynamics* 16, 5-21.

Jones, Steve (2000), *Darwin's Ghost. The Origin of Species Updated*, New York: Ballantine Books.

Kessler, Brett (2001), *The Significance of Word Lists*, Stanford: Center for the Study of Language and Information.

Inglehart, Ronald and Hans-Dieter Klingemann (2000), Genes, culture, democracy, and hapiness, in E. Diener and E.M. Suh, *Subjective Well-Being Across Cultures*, Cambridge, MA: MIT Press.

Le Breton, Michel, Ignacio Ortuno-Ortin and Shlomo Weber (2003), Nation formation and genetic diversity, manuscript.

McFadyen, Stuart, Colin Hoskins and Adam Finn (2004), Measuring the cultural discount in the price of exported U.S. television programs, in Victor Ginsburgh, ed., *Economics of Culture*, Amsterdam: Elsevier.

Rask, Morten (2004), Literature about cultural distances on the Net, http://www.morton-rask.dk/2001f.htm.

Ruhlen, Merritt (1994), *The Origin of Language*, New York: John Wiley and Sons.

Swadesh, Morris (1952), Lexicostatistic dating of prehistoric ethnic contacts, *Proceedings of the American Philosophical Society*, vol. 96, 452-463.

Table 1
The Dyen Matrix of Linguistic Distances
Between Some Languages

	English	French	German	Italian	Spanish
Danish	0.407	0.759	0.293	0.737	0.750
Dutch	0.392	0.756	0.162	0.740	0.742
English	0	0.764	0.422	0.753	0.760
Finnish	1.000	1.000	1.000	1.000	1.000
French	0.764	0	0.756	0.197	0.291
German	0.422	0.756	0	0.735	0.747
Greek	0.838	0.843	0.812	0.822	0.833
Italian	0.753	0.197	0.735	0	0.212
Portuguese	0.760	0.291	0.753	0.227	0.126
Spanish	0.760	0.291	0.747	0.212	0
Swedish	0.411	0.756	0.305	0.741	0.747

Notes. Since Finnish is not a Indo-European language, it is not included in Dyen et al (1992). Given its linguistic remoteness, its distance to every language in the table was set to 1. Source: Dyen, Kruskal and Black (1992, pp. 102-117).

Table 2
Languages: Estimation Results

	Coefficient	Standard error
Population speaking language i (α_1)	-0.058	0.069
Population speaking language j (α_2)	0.625*	0.057
Distance between languages i and j (α_3)	-0.954*	0.200
Intercept (α_{0E})	0.000	-
Intercept (α_{0F})	-0.112	0.062
Intercept (α_{0G})	-0.233*	0.061
Intercept (α_{0S})	-0.514*	0.050
R-square	0.758	
No. of observations	46	

Notes. A * indicates that the coefficient is significantly different from zero at the 5 or 1 percent probability level. The number of observations is 12 for French and Spanish, since these are foreign languages in 12 of the 13 countries. The number is 11 for English (spoken in the UK and Ireland) and German (spoken in Austria and Germany). Source: Ginsburgh, Ortuno-Ortin and Weber (2004b).

Table 3
Genetic Distances Between Some Populations

	Basque	Spanish	Sardinian	Italian	Scottish	English	Lapp
Basque	0	104	261	141	146	119	629
Spanish	104	0	295	61	100	47	452
Sardinian	261	295	0	221	112	51	339
Italian	141	61	221	0	112	51	339
Scottish	146	100	357	112	0	27	447
English	119	47	340	51	27	0	404
Lapp	629	452	667	339	447	404	0

Actual calculated distances are multiplied by 10,000. Source: Cavalli-Sforza et al. (1994, p. 270).

Table 4
Cultural Distances Between the United States and Selected Countries

	Power	Individualism	Masculinity	Uncertainty Avoidance	Confucian Dynamism
Australia	36	90	61	51	31
Austria	11	55	79	70	n.a.
Belgium	65	75	54	94	n.a.
Canada	39	80	52	48	23
Denmark	18	74	16	23	n.a.
Finland	33	63	26	59	n.a.
France	68	71	43	86	n.a.
Germany	35	67	66	65	31
Great Britain	35	89	66	35	25
Greece	60	35	57	112	n.a.
Ireland	28	70	68	35	n.a.
Israel	13	54	47	81	n.a.
Italy	50	76	70	75	n.a.
Netherlands	38	80	14	53	44
Norway	31	69	8	50	n.a.
Portugal	63	27	31	104	n.a.
Spain	57	51	42	86	n.a.
Sweden	31	71	5	29	33
Switzerland	34	68	70	58	n.a.
Turkey	66	37	45	85	n.a.
Yugoslavia	76	27	21	88	n.a.

Source: McFadyen, Hoskins and Finn (2004). All the data are available on Hofstede's website http://geert-hofstede.international-business-center.com/index.shtml

Table 5
Correlations Between Linguistic and Cultural Distances

	Lang.	Power	Individ.	Masc.	Uncert. Avoid.
Language	1				
Power	0.205	1			
Individualism	0.254	0.111	1		
Masculinity	-0.092	0.031	-0.128	1	
Uncertainty Avoidance	0.319	0.567	0.404	0.083	1

Source: Ginsburgh and Noury (2004).

Table 6
The Eurovision Song Contest: Estimation Results

	(1)	(2)	(3)	(4)
Quality	0.912**	0.914**	0.908**	1.034**
	(0.024)	(0.025)	(0.025)	(0.066)
Lagged vote	0.027*	0.024	0.015	0.016
	(0.013)	(0.013)	(0.014)	(0.013)
Order of perf.	0.003	0.002	0.003	-0.002
	(0.009)	(0.008)	(0.008)	(0.007)
Host country	0.175	0.188	0.178	-
	(0.236)	(0.235)	(0.237)	
Sung in English	0.142	0.181	0.137	-
	(0.135)	(0.135)	(0.136)	
Sung in French	0.347*	0.340*	0.328	-
	(0.167)	(0.167)	(0.172)	
Male singer	0.140	0.146	0.152	-
	(0.127)	(0.128)	(0.127)	
Duet	0.226	0.156	0.189	-
	(0.198)	(0.198)	(0.197)	
Group	0.101	0.082	0.078	-
	(0.131)	(0.130)	(0.130)	
Language	-	-0.964**	-0.595**	-0.642**
		(0.194)	(0.212)	(0.205)
Power	-	-	-0.890*	-0.603
			(0.422)	(0.457)
Individualism	-	-	-0.084	0.015
			(0.400)	(0.038)
Masculinity	-	-	-0.487	-0.459
			(0.257)	(0.256)
Uncertainty Avoidance	-	-	-0.909**	-1.032**
			(0.320)	(0.335)
Intercept	0.058	0.615**	1.087**	0.884**
	(0.128)	(0.173)	(0.209)	(0.275)
R-square	0.30	0.30	0.31	0.30
No. of obs.	4,074	4,074	4,074	4,074

Robust standard errors appear between brackets. ** and * for sigificantly different from zero at the 1 and 5 percent probability level. Source: Ginsburgh and Noury (2004).