

ONLINE APPENDIX

Ancient Origins of the Global Variation in Economic Preferences (AEA P&P 2020)

by Anke Becker, Benjamin Enke and Armin Falk

1 Preferences Data

The data on risk, time, and social preferences are part of the Global Preference Survey (GPS), a recently released survey dataset on economic preferences from representative population samples in 76 countries that is described in detail in [Falk et al. \(2018\)](#).¹ The GPS data (i) are based on an ex-ante experimental validation procedure of the survey items; (ii) make use of representative samples in 76 countries, for a total sample size of 80,000 participants worldwide; (iii) are geographically representative in that they cover countries on all continents and of all development levels; and (iv) were collected through the professional infrastructure of the Gallup World Poll (see [Falk et al., 2018](#), for details).

Risk Taking. The set of survey items includes two measures of the underlying risk preference – one qualitative subjective self-assessment and one quantitative measure. The subjective self-assessment directly asks for an individual’s willingness to take risks: *“Generally speaking, are you a person who is willing to take risks, or are you not willing to do so? (0–10)”*

The quantitative measure is derived from a series of five interdependent hypothetical binary lottery choices. In each of the five questions, participants had to decide between a 50-50 lottery to win x or nothing (which was the same in each question) and varying safe payments y . The questions were interdependent in the sense that the choice of a lottery resulted in an increase of the safe amount y being offered in the next question, and conversely. By adjusting the safe payment according to previous choices, the questions “zoom in” around the respondent’s certainty equivalent. The self-assessment and the outcome of the quantitative lottery procedure were then aggregated into a single index which describes an individual’s degree of risk taking.

Patience. The measure of patience is also derived from the combination of responses to two survey measures, one with a quantitative and one with a qualitative format. The quantitative survey measure consists of a series of five hypothetical binary choices

¹See <https://www.briq-institute.org/global-preferences/home>

between immediate and delayed financial rewards. Similar to the elicitation of risk preferences, the questions were interdependent in the sense that the delayed payment was increased or decreased depending on previous choices. The qualitative measure of patience is given by the respondent's self-assessment regarding their willingness to wait on an 11-point Likert scale, asking "how willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?".

Prosociality: Altruism, Positive Reciprocity, and Trust. The GPS includes six survey items which map into three prosocial traits: altruism, positive reciprocity, and trust. While these behavioral traits are conceptually distinct, they share in common that they are commonly associated with "positive" social interactions.

Altruism was measured through a combination of one qualitative and one quantitative item, both of which are related to donation. The qualitative question asked people how willing they are to give to good causes without expecting anything in return on an 11-point scale. The quantitative scenario depicted a situation in which the respondent unexpectedly received € 1,000 and asked them to state how much of this amount they would donate.

People's propensity to act in a positively reciprocal way was also measured using one qualitative item and one question with a quantitative component. First, respondents were asked to provide a self-assessment about how willing they are to return a favor on an 11-point Likert scale. Second, participants were presented a choice scenario in which they were asked to imagine that they got lost in an unfamiliar area and that a stranger – when asked for directions – offered to take them to their destination. Participants were then asked which out of six presents they would give to the stranger as a "thank you." Finally, to measure trust, people were asked whether they assume that other people only have the best intentions (Likert scale, 0-10).

Because these three variables are highly correlated and to reduce the number of dependent variables (and associated multiple testing concerns), we collapse them into a prosociality score that consists of the unweighted average of the three variables.

Negative Reciprocity. Negative reciprocity was elicited through three self-assessments. First, people were asked how willing they are to take revenge if they are treated very unjustly, even if doing so comes at a cost (0-10). The second and third item probed respondents about their willingness to punish someone for unfair behavior, either towards *themselves* or towards a *third person*.

As discussed in [Falk et al. \(2018\)](#), the preference measures are constructed by linearly combining responses to the survey items using weights that are derived from the experimental validation procedure.

We investigate the origins of this heterogeneity through a bilateral regression approach in which absolute differences in preferences serve as dependent variable. Thus, we compute the absolute difference in a given trait. Furthermore, for each population pair, we calculate an overall summary statistic of preference differences by summing up these absolute differences across preference dimensions. This summary statistic can be understood as a measure of overall preference dissimilarity.

2 Data on Ancestral Distance

2.1 Data on Genetic Distance

Whenever populations break apart, they stop interbreeding, thereby preventing a mixture of the respective genetic pools. However, since every genetic pool is subject to random drift or local selection pressures, geographical separation implies that over time the genetic distance between sub-populations gradually becomes (on average) larger. Thus, the genealogical relatedness between two populations reflects the length of time elapsed since these populations shared common ancestors. Our country-level data on the so-called F_{ST} genetic distance are taken directly from [Spolaore and Wacziarg \(2009\)](#) and [Spolaore and Wacziarg \(2017\)](#). These are two different and independent datasets on genetic distance. The first is derived from data on classic genetic markers, while the second is based on microsatellite variation. In both cases, Spolaore and Wacziarg match the pairwise ethnic-group level distances to countries and then compute the expected genetic distance between two countries. These genetic distance measures have the same conceptual basis, but are based on different biological information and samples.

Technically, genetic distance constitutes an index of expected heterozygosity, which can be thought of as the probability that two randomly matched individuals will be genetically different from each other in terms of a pre-defined spectrum of genes. Indices of heterozygosity are derived using data on allelic frequencies, where an allele is a particular variant taken by a gene. Intuitively, the relative frequency of alleles at a given locus can be compared across populations and the deviation in frequencies can then be averaged over loci.

This is the approach pursued in the work of the population geneticists [Cavalli-Sforza, Menozzi and Piazza \(1994\)](#). The main dataset assembled by these researchers consists of data on 128 different alleles for 42 world populations. By aggregating differences in these allelic frequencies, the authors compute the F_{ST} genetic distance, which provides a comprehensive measure of genetic relatedness between any pair of 42 world populations. Since genetic distances are available only at the population rather than

at the country level, [Spolaore and Wacziarg \(2009\)](#) matched the 42 populations in [Cavalli-Sforza, Menozzi and Piazza \(1994\)](#) to countries.² Thus, the genetic distance measures we use measure the expected genetic distance between two randomly drawn individuals, one from each country, according to the contemporary composition of the population.

Recently, [Spolaore and Wacziarg \(2017\)](#) introduced a new dataset of cross-country F_{ST} genetic distances that is based on the work by [Pemberton, DeGiorgio and Rosenberg \(2013\)](#). While the data from [Cavalli-Sforza, Menozzi and Piazza \(1994\)](#) are based on classic genetic markers, this new dataset is based on microsatellite variation, covering 645 microsatellite loci and 267 populations, thus providing a more comprehensive and detailed coverage of world populations. [Spolaore and Wacziarg \(2017\)](#) again matched these population-level F_{ST} distances to countries using ethnic composition data from [Fearon \(2003\)](#). In sum, this more recent genetic distance measure has the same conceptual basis, but is based on different biological information and samples.

2.2 Data on Linguistic Distance

Population break-ups produce not only diverging gene pools, but also differential languages. The construction of linguistic distance follows the methodology proposed by [Fearon \(2003\)](#). The Ethnologue project classifies all languages of the world into language families, sub-families, sub-sub-families etc., which gives rise to a language tree. In such a tree, the degree of relatedness between different languages can be quantified as the number of common nodes two languages share. For each country pair, we calculate the weighted linguistic distance according to the population shares speaking a particular language in the respective countries today.

To compute the linguistic distance between any two languages, we apply the following procedure. If two languages belong to different language families, the number of common nodes is 0. In contrast, if two languages are identical, the number of common nodes is 15. Following [Fearon \(2003\)](#), who argues that the marginal increase in the degree of linguistic relatedness is decreasing in the number of common nodes, we

²To this end, the authors used ethnic composition data from [Fearon \(2003\)](#): the data by [Cavalli-Sforza, Menozzi and Piazza \(1994\)](#) contain information on the groups that were sampled to obtain genetic distance estimates, and these groups can be matched one-to-one to the ethnic groups that populate countries. Thus, the data from one group in [Cavalli-Sforza, Menozzi and Piazza \(1994\)](#) can be assigned to sub-populations in potentially multiple countries, so that, in principle, even the relatively small number of 42 populations is sufficient to compute genetic distances between more than 100 countries.

transformed these data according to

$$\text{Linguistic distance (tree)} = 1 - \sqrt{\frac{\# \text{ Common nodes}}{15}}$$

to produce distance estimates between languages in the interval $[0, 1]$. We restricted the Ethnologue data to languages which either make up at least 5% of the population in a given country, or are an interview language in the GPS.

To convert these language distances into distances at the country level, we compute the expected linguistic distance between two randomly selected individuals, one from each country. Formally, suppose there are N languages. Let $s_{1,i}$ be the share of the population in country 1 which speaks language i and denote by $d_{i,j}$ the linguistic distance between languages i and j . Then, the (weighted) linguistic distance between countries 1 and 2 is given by

$$\text{Linguistic distance}_{1,2} = \sum_{i=1}^N \sum_{j=1}^N (s_{1,i} \times s_{2,j} \times d_{i,j})$$

As a second and complementary measure of linguistic distance, we use a lexicostatistical measure of linguistic distance developed as part of the Automatic Similarity Judgment Program (ASJP) at the Max Planck Institute for Evolutionary Anthropology (Wichmann, Holman and Brown, 2016). This measure has been developed partly to allow for analyses of when languages diverged from each other (Holman et al., 2011). The measure is based on a list of 40 words with universal meaning across languages (e.g., “I”, “hand”, and “night”). The measure of linguistic distance is constructed by counting the number of phonetic edits needed to rewrite each word from one language spelling to another. The ASJP database contains the full matrix of linguistic distances between more than 4,500 languages. We again convert these language-level distances into country-level distances by calculating the weighted linguistic distance according to the population shares speaking a particular language in the respective countries today.

2.3 Construction of Composite Measure of Ancestral Distance.

In sum, we have access to four proxies for ancestral distance (or temporal distance). Given that these measures follow different methods of construction and are likely to be plagued by measurement error, we construct a composite index of ancestral distance as unweighted average of the standardized values (z-scores) of the four distance variables. We standardize the ancestral distance measure into a z-score to ease interpretation of regression coefficients. To get a sense for the underlying variation, note that a one

standard deviation increase in ancestral distance corresponds to moving from a country pair such as Finland and Portugal to a country pair such as Bolivia and Philippines or Poland and Tanzania.

3 Formal Framework

3.1 Model

We here formally illustrate how both of the channels discussed in the main text (historical experiences and genetic pools) yield the prediction that longer separation implies larger absolute differences in preferences. The framework builds on [Spolaore and Wacziarg \(2009\)](#).

We conceptualize both experiences and genetic changes through population-specific stochastic shocks. We then show that these shocks “add up” over time and hence generate a relationship between length of separation and preference differences. Importantly, neither the framework nor our empirical exercise distinguishes (or is even intended to distinguish) between genetic and experience-based mechanisms. Given recent evidence on gene-environment interactions (see [Manuck and McCaffery, 2014](#), for an overview), the long-run focus of our analysis renders such a distinction fundamentally misguided.

A seemingly important assumption is how the population-specific shocks are distributed across populations and time. Evidently, making intuitively appealing assumptions such as “populations that have been separated for a shorter time and hence likely live close geographically are subject to more similar shocks”, would trivially yield the prediction that temporal distance predicts preference differences. However, we derive our prediction in its arguably starkest form by showing that preference differences should depend on temporal difference *even* if the shocks are independently distributed across time and space.

Formally, suppose that there is a set of N contemporary populations. In period $t = 0, 1, \dots, T$, each population i has a scalar-representable preference endowment x_i^t . In period $t = 0$, all contemporary populations were part of one “parental” population and we normalize the preference endowment to $x^0 = 0$. Over time, populations successively broke apart from each other. For each time $t = 0, 1, \dots$ let \mathcal{P}_t be a partition of $\{1, \dots, N\}$, that is, \mathcal{P}_t is a collection of disjoint nonempty sets whose union is $\{1, \dots, N\}$. The elements of \mathcal{P}_t represent the different populations at time t . For each $t \geq 0$ and $i \in \{1, \dots, N\}$ let $P_t(i)$ be the unique $A \in \mathcal{P}_t$ that contains i .

In each period, a given population’s preference endowment is subject to a random shock, which could result from experiences or changes in the genetic pool, or both. That is, as long as two populations are not separated, they get hit by the same shock, but once they split up, they are subject to separate, and potentially different, shocks. For each $t \geq 1$ and each $A \in \mathcal{P}_t$ let ϵ_A^t be such a random shock. Even though this is technically

redundant, we will assume that the shocks have mean zero to ease interpretation. Let

$$x_i^t = \sum_{\tau=1}^t \epsilon_{P_\tau(i)}^\tau.$$

That is, a population's preference endowment in period t is given by the sum of the accumulated shocks. The object of interest in the empirical analysis is the expression

$$E \left[\left| x_i^T - x_j^T \right| \right]$$

for $i, j \in \{1, \dots, N\}$. We will show that under arguably very mild assumptions this absolute difference in preferences between populations i and j is increasing in the number of periods in which the populations were separated. Fix $T \geq 1$. For populations $i, j \in \{1, \dots, N\}$ let $s_{ij} = |\{t \in \{1, \dots, T\} : P_t(i) \neq P_t(j)\}|$. Thus, s_{ij} is the number of periods up to time T where i and j were separated.

To derive our main prediction, we will assume that the preference shocks are independently and identically distributed across time and populations. As noted above, this assumption *only* serves to derive the prediction in its starkest (and arguably non-trivial) form. As we discuss below, other assumptions would often trivially generate the prediction that longer separation induces larger preference differences.

Proposition 1. *Suppose the shocks ϵ_A^t , $A \in \mathcal{P}_t$, $t = 1, \dots, T$, are i.i.d. nondegenerate integrable random variables. Let $i, j, k, l \in \{1, \dots, N\}$. Then*

$$s_{ij} > s_{kl} \quad \Leftrightarrow \quad E \left[\left| x_i^T - x_j^T \right| \right] > E \left[\left| x_k^T - x_l^T \right| \right].$$

The proof is below.³ To see the basic intuition, suppose that populations i and j are still one population in T , i.e., they got hit by the same sequence of shocks, so that their absolute difference in preferences is zero. Suppose further that populations i and k were separated for one period, implying that their absolute difference in preferences is given by $|x_i^T - x_k^T| = |\epsilon_i - \epsilon_k|$. In expectation, this expression is strictly greater than zero. The proposition shows that this intuition holds for arbitrary population breakups and time spans. Hence, we state the following testable hypothesis:

Hypothesis. *The absolute difference in preferences between two populations increases in their length of separation.*

³We are deeply indebted to Lorens Imhof for proposing the proof to us.

Note that the assumptions in Proposition 1 are sufficient, but not necessary, to generate the prediction that longer separation implies larger expected absolute differences.

Remark 1. *It is conceivable that the preference shocks are drawn from different distributions along the migratory path, say because the further populations migrate the larger the average preference shock. However, if preferences evolved monotonically along the migratory path, then temporal distance trivially ought to be predictive of preference differences, which is why we refrain from making such strong assumptions. In addition, there is no biological principle according to which the evolution of a scalar-representable trait must follow a monotonic path. While there are reasons to believe that traits like risk aversion, time preference, or altruism are subject to local selection pressures, these selection pressures might operate in different directions along the migratory path as groups of humans and their descendants pass through many different environments.*

Remark 2. *The assumption that preference shocks are independent of each other across space is likely to be unrealistic. However, again, making natural assumptions on the dependence of the shocks across populations would trivially imply the prediction that populations with low temporal distance have more similar preference profiles.*

3.2 Proofs

We are deeply indebted to Lorens Imhof for proposing these proofs to us.

Proof of Proposition 1. We have

$$x_i^T - x_j^T = \sum_{t=1}^T \epsilon_{P_t(i)}^t - \sum_{t=1}^T \epsilon_{P_t(j)}^t = \sum_{\substack{t=1, \dots, T, \\ P_t(i) \neq P_t(j)}} (\epsilon_{P_t(i)}^t - \epsilon_{P_t(j)}^t),$$

which is a sum of s_{ij} differences of shocks. Let $u_1, \dots, u_T, v_1, \dots, v_T$ be i.i.d. random variables having the same distribution as the ϵ_A^t . Then $x_i^T - x_j^T$ has the same distribution as $\sum_{n=1}^{s_{ij}} (u_n - v_n)$. A similar argument shows that $x_k^T - x_l^T$ has the same distribution as $\sum_{n=1}^{s_{kl}} (u_n - v_n)$. In particular,

$$E \left[\left| x_i^T - x_j^T \right| \right] = E \left[\left| \sum_{n=1}^{s_{ij}} (u_n - v_n) \right| \right]$$

and

$$E \left[\left| x_k^T - x_l^T \right| \right] = E \left[\left| \sum_{n=1}^{s_{kl}} (u_n - v_n) \right| \right].$$

The claimed equivalence will follow if we can show that

$$E\left[\left|\sum_{n=1}^m(u_n - v_n)\right|\right] < E\left[\left|\sum_{n=1}^{m+1}(u_n - v_n)\right|\right], \quad m = 0, \dots, T - 1. \quad (1)$$

We will apply Lemma 1 below. Fix $m \in \{0, \dots, T - 1\}$ and let $y = \sum_{n=1}^m(u_n - v_n)$ and $z = u_{m+1} - v_{m+1}$. Then y and z are independent integrable random variables. Moreover, $E[z] = E[u_{m+1}] - E[v_{m+1}] = 0$ and since the shocks are nondegenerate,

$$\begin{aligned} P(z \neq 0) &\geq P(u_{m+1} > E[u_{m+1}], v_{m+1} < E[v_{m+1}]) \\ &= P(u_{m+1} > E[u_{m+1}])P(v_{m+1} < E[v_{m+1}]) > 0. \end{aligned}$$

Finally, for every $c > 0$, there exists $\xi \in \mathbb{R}$ such that $P(|\sum_{n=1}^m u_n - \xi| < \frac{c}{2}) > 0$. Hence,

$$\begin{aligned} P(|y| < c) &\geq P\left(\left|\sum_{n=1}^m u_n - \xi\right| < \frac{c}{2}, \left|\sum_{n=1}^m v_n - \xi\right| < \frac{c}{2}\right) \\ &= P\left(\left|\sum_{n=1}^m u_n - \xi\right| < \frac{c}{2}\right)^2 > 0, \end{aligned}$$

which shows that the support of the distribution of y contains the point 0. Inequality (1) now follows from Lemma 1. \square

Lemma 1. *Let y and z be independent integrable random variables. Suppose that 0 is in the support of the distribution of y , $E[z] = 0$ and $P(z \neq 0) > 0$. Then $E[|y + z|] > E[|y|]$.*

Proof. Since y and z are independent, $E[z|y] = E[z] = 0$, and so

$$E[|y + z||y] \geq |E[y + z|y]| = |E[y|y]| = |y|. \quad (3)$$

Using the inequality $|y + z| \geq |z| - |y|$ and again the independence of y and z , we obtain

$$E[|y + z||y] \geq E[|z||y] - E[|y||y] = E[|z|] - |y|.$$

Hence, on the event $\{2|y| < E[|z|]\}$,

$$E[|y + z||y] > |y|.$$

The assumption that $P(z \neq 0) > 0$ implies that $E[|z|] > 0$, and since 0 is contained in the support of the distribution of y , $P(2|y| < E[|z|]) > 0$. That is, inequality (2) holds almost everywhere and the inequality is strict on a set of positive probability. Taking expectations we get $E[|y + z|] > E[|y|]$. \square

References

- Cavalli-Sforza, Luigi, Paolo Menozzi, and Alberto Piazza.** 1994. *The History and Geography of Human Genes*. Princeton University Press.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde.** 2018. "Global evidence on economic preferences." *The Quarterly Journal of Economics*, 133(4): 1645–1692.
- Fearon, James D.** 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, 8(2): 195–222.
- Holman, Eric W., Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown, et al.** 2011. "Automated Dating of the World's Language Families Based on Lexical Similarity." *Current Anthropology*, 52(6): 841–875.
- Manuck, Stephen B., and Jeanne M. McCaffery.** 2014. "Gene-Environment Interaction." *Annual Review of Psychology*, 65: 41–70.
- Pemberton, Trevor J., Michael DeGiorgio, and Noah A. Rosenberg.** 2013. "Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation." *G3: Genes | Genomes | Genetics*, 3: 891–907.
- Spolaore, Enrico, and Romain Wacziarg.** 2009. "The Diffusion of Development." *Quarterly Journal of Economics*, 124(2): 469–529.
- Spolaore, Enrico, and Romain Wacziarg.** 2017. "Ancestry and Development: New Evidence." *Working Paper*.
- Wichmann, Søren, Eric W. Holman, and Cecil H. Brown.** 2016. "The ASJP Database (version 17)." URL: <http://email.eva.mpg.de/~wichmann/ASJPHomePage.htm>, 3.