NBER WORKING PAPER SERIES

ANCIENT ORIGINS OF THE GLOBAL VARIATION IN ECONOMIC PREFERENCES

Anke Becker
Benjamin Enke
Armin Falk

Working Paper 24291
http://www.nber.org/papers/w24291

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2018

An earlier version of this paper circulated as “The Ancient Origins of Cross-Country Heterogeneity in Risk Preferences”. We thank Markus Antony for extraordinary help in collecting the GPS data and Ammar Mahran and Patricia Sun for outstanding research assistance. We also thank Quamrul Ashraf and Oded Galor for data sharing. Armin Falk acknowledges financial support from the European Research Council through ERC #209214. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Ancient Origins of the Global Variation in Economic Preferences
Anke Becker, Benjamin Enke, and Armin Falk
NBER Working Paper No. 24291
February 2018
JEL No. D03

ABSTRACT

Variation in economic preferences is systematically related to both individual and aggregate economic outcomes, yet little is known about the origins of the worldwide preference variation. This paper uses globally representative data on risk aversion, time preference, altruism, positive reciprocity, negative reciprocity, and trust to uncover that contemporary preference heterogeneity has its roots in the structure of the temporally distant migration patterns of our very early ancestors: In dyadic regressions, differences in preferences between populations are significantly increasing in the length of time elapsed since the ancestors of the respective groups broke apart from each other. To document this pattern, we link genetic and linguistic distance measures to population-level preference differences (i) in a wide range of cross-country regressions, (ii) in within-country analyses across groups of migrants, and (iii) in analyses that leverage variation across linguistic groups. While temporal distance drives differences in all preferences, the patterns are strongest for risk aversion and prosocial traits.

Anke Becker
Harvard University
Department of Human Evolutionary Biology
11 Divinity Ave
Cambridge MA 02138
ankebecker@fas.harvard.edu

Benjamin Enke
Department of Economics
Harvard University
Littauer M008
Cambridge, MA 02138
and NBER
enke@fas.harvard.edu

Armin Falk
Department of Economics
University of Bonn
Adenauerallee 24-42, 53113 Bonn
armin.falk@uni-bonn.de

1 Introduction

Risk, time, and social preferences form the building blocks of a large class of models in both micro- and macroeconomics. Empirical work shows that these preferences vary substantially within populations and – in line with economic models – predict a large set of individual-level economic decisions ranging from stock and labor market behavior over savings and schooling choices to prosocial activities (e.g., Borghans and Golsteyn, 2006; Sutter et al., 2013; Kosfeld and Rustagi, 2015; Cohn and Maréchal, forthcoming). Recent work has documented that preferences exhibit large variation not just between individuals, but also across countries (Herrmann et al., 2008; Gächter and Schulz, 2016). Using the newly constructed Global Preference Survey, Falk et al. (forthcoming) document that this cross-country variation in preferences predicts important economic outcomes, including per capita income, entrepreneurial activities, donations and volunteering, and the frequency of armed conflicts. The insights that preferences exhibit large variation both across and within countries, and that this heterogeneity correlates with economic outcomes at both levels of aggregation, raise the question of the origins of large-scale preference variation.

With some exceptions that we discuss below, work on endogenous preferences thus far has mostly focused on the role of *contemporary* environmental conditions. This line of research has documented that risk, time, and social preferences respond in predictable ways to educational interventions (Alan and Ertac, Forthcoming; Cappelen et al., 2016), circles of friends and mentors (Kosse et al., 2016; Rao, 2018), conflict (Voors et al., 2012), or income shocks (Haushofer et al., 2013).

This paper takes this literature a step further through a complementary approach in which we (i) focus on explaining the global variation in preferences, as opposed to heterogeneity within a given population, (ii) investigate the *very* deep origins of preference heterogeneity, and (iii) consider multiple preferences and attitudes in a unified empirical approach. The key idea is to link the global preference heterogeneity to the structure of mankind's ancient migration out of Africa, a sequence of events that has attracted much recent interest in the comparative development literature (Spolaore and Wacziarg, 2009, 2017; Ashraf and Galor, 2013; Ashraf et al., 2014). Building on a simple dynamic model of preference formation, we document that these temporally distant migration movements have shaped today's heterogeneity in risk, time, and social preferences, both across and within countries, albeit to heterogeneous degrees across preferences. To our knowledge, this is the first systematic evidence that links historical migratory movements to the structure of economic preferences, and the first attempt to explain some of the between country heterogeneity that the Global Preference Survey has brought to light.

According to the widely accepted "Out of Africa hypothesis" of human development, starting around 50,000-60,000 years ago, early mankind migrated out of East Africa and continued to explore and populate our planet through a series of successive migratory steps that are referred to as the "great human expansion" (see Henn et al., 2012, for an overview). Each of these steps consisted of some sub-population breaking apart from the previous colony and moving on to found new settlements. This pattern implies that some contemporary population pairs have spent a longer time of human history apart from each other than others. As a result, the time elapsed since two groups shared common ancestors differs across today's population pairs. The key idea underlying our analysis is that these differential time frames of separation might have generated heterogeneity in preferences over risk, time, and social interactions. First, populations that have spent a long time apart from each other were exposed to different historical experiences and environments, which could affect preferences. Second, due to random genetic drift or local selection pressures, long periods of separation lead to different population-level genetic endowments, which might in turn shape attitudes. The paper develops a model of preference evolution in the presence of stochastic shocks to show that both the genetic and the experience-based channel imply the prediction that populations that have been separated for a longer time should exhibit larger (absolute) differences in preferences.

We investigate this hypothesis at three levels of analysis, i.e., across countries, within countries across populations of different ancestry, and across linguistic groups. For this purpose, the analysis combines data on economic preferences around the globe with proxies for long-run human migration patterns and the resulting temporal distances. Our data on preferences stem from the Global Preference Survey (GPS), which includes experimentally validated survey measures of risk, time, and social preferences of 80,000 people from a geographically and economically diverse set of 76 countries (Falk et al., forthcoming). These data allow the computation of nationally representative levels of risk aversion, patience, altruism, positive reciprocity, negative reciprocity, and trust.

The analysis combines these data with proxies for the temporal patterns of ancient population fissions, i.e., proxies for the length of time since two populations shared common ancestors. First, we employ the F_{ST} genetic distance between populations (Spolaore and Wacziarg, 2009, 2017). As population geneticists have long noted, whenever two populations split apart from each other in order to found separate settlements, their genetic distance increases over time due to random genetic drift. Thus, the genetic distance between two populations is a measure of *temporal distance*. Second, we make use of the observation that linguistic differences closely track the structure of separation of human populations, and employ two measures of linguistic distance as explanatory variable. We collapse these measures into a summary statistic of temporal distance between populations. The origins of the vast majority of the variation in this temporal distance measure date back thousands of years.

The empirical analysis of the relationship between preferences and ancient migration patterns starts with documenting that the absolute difference in risk, time, and social preferences between two countries is significantly increasing in the respective populations' temporal distance. In quantitative terms, an increase of one standard deviation in temporal distance is associated with an increase of about 22% of a standard deviation of differences in preferences, which is a larger effect than the corresponding correlation between differences in per capita income and preference differences. An array of robustness checks establishes that our cross-country results are robust to employing the genetic, migratory, and linguistic distance variables separately. These results are strongest for risk aversion and the prosocial traits altruism, positive reciprocity, and trust; similar, but weaker, findings hold for patience and negative reciprocity. We further document that the results are robust to accounting for differences in individual-level observables across countries and to taking into account potential culture-dependent interpretations of the survey items.

"Controlling" for contemporary environmental conditions is not necessarily meaningful in our context. After all, those very conditions could often represent the mechanisms underlying the effect of temporal distance on preference differences, in particular when geographic and climatic variables capture highly persistent variation across populations. Nonetheless, the analysis addresses potential concerns that only variation in contemporary environments generates the results through a set of conditional regressions. The results establish that the effect of temporal distance on differences in risk aversion and prosocial traits is robust to an extensive set of covariates, including controls for differences in the countries' demographic composition, their geographic position, geographic distance metrics, prevailing climatic and agricultural conditions, institutions, and economic development. In all of the corresponding regressions, the point estimate is very stable, which suggests that unobserved heterogeneity is unlikely to drive our results (Altonji et al., 2005). In contrast, the relationships between patience and negative reciprocity on the one hand and temporal distance on the other hand disappear once covariates are accounted for. Still, a perhaps interesting insight of our analysis is that temporal distance is much more predictive of preference differences than simple geographic distance measures such as geodesic distance. This suggests that it is indeed the structure of historical population movements rather than simple shortest-distance calculations that explain heterogeneity in preferences.

In a next step, the paper studies the subnational relationship between preferences and temporal distance. For this purpose, the analysis uses information on individuals' country of birth in the GPS. For each country of residence, we construct virtual populations by averaging preferences across migrants from a given country of birth (Fernández and Fogli, 2006; Giuliano, 2007), and then assign a within-country population pair the temporal distance of the respective countries of birth. In essence, the resulting regressions compare, say, the difference in preferences between Italians and Turks with that of Chinese and Norwegians, all of whom currently reside in Germany. These regressions include both country of birth and country of residence fixed effects, i.e., leverage variation in preferences and temporal distance, while holding the current location constant across populations.

The results of the within-country exercise are even stronger than those established in the cross-country case. Across all preferences, temporal distance is strongly predictive of preference variation. Again, these results survive a number of robustness checks including employing each temporal distance proxy separately, or accounting for differences in individual-level observables. This set of results not only adds credibility to our identification strategy, but also represents a methodological innovation on past work on temporal or genetic distance, which has exclusively considered cross-country variation.

In the cross-country and within-country analyses, the basic unit of observation is a population as defined by country of residence or country of birth. An alternative way of conceptualizing a population is through linguistic groups. After all, from an evolutionary perspective, populations that speak the same language – even if they reside in different countries such as Germans and Austrians – might (approximately) be considered one population. Generalizing this logic, the analysis exploits information on respondents' interview language in the GPS to compute average preferences at the language group level, and to relate the difference in preferences between these groups to their linguistic distance as proxy for temporal distance. The results document that linguistic distance is significantly related to differences in preferences, (i) in simple baseline regressions, (ii) when individual-level differences in observables are accounted for, and (iii) when colonial languages – for which the overlap between language and temporal distance is small – are excluded from the analysis. Thus, the relationship between preference differences and temporal distance does not hinge on defining populations through their country of residence or birth.

Taken together, the results establish that – across countries of residence, across countries of birth within countries of residence, and across linguistic groups – the longer two populations have been separated in the course of human history, the more they differ in terms of their economic preferences. These effects are smaller and less robust for patience and negative reciprocity, which suggests that these preferences may have more recent origins. This insight is potentially useful for theories on the evolution of preferences (Bisin and Verdier, 2001; Doepke and Zilibotti, 2013, 2017).[1]

Our paper is related to other work that uses genetic or linguistic data, albeit in contexts other than economic preferences (Spolaore and Wacziarg, 2009, 2016; Ashraf and Galor, 2013; Ashraf et al., 2014; Desmet et al., 2011; Chen, 2013; Özak, 2016).[2] Our results lend support to the idea that pairwise (genetic) distance captures cultural differences, as is implicitly or explicitly assumed in reducd-form analyses of the relationship between temporal distance and development (Spolaore and Wacziarg, 2009, 2017).

By uncovering that population-level preference profiles are endogenous to temporal distance, we also contribute to the line of work on the historical origins of economically relevant attitudes (Guiso et al., 2009; Durante, 2009; Voigtländer and Voth, 2012; Alesina et al., 2013; Galor and Özak, 2016; Olsson and Paik, 2016; Enke, 2017; Becker, 2017). This work has focused on traits that are conceptually distinct from the tightly measured preference parameters in the GPS.

The remainder of this paper proceeds as follows. In Section 2, we develop our hypothesis on the relationship between the structure of migratory movements and preferences, while Section 3 presents the data. Section 4 discusses the cross-country evidence, Section 5 the within-country results, and Section 6 the analysis across language groups, respectively. Section 7 concludes.

2 Preferences and the Great Human Expansion

According to the widely accepted "Out of Africa" theory of the origins and the dispersal of early humans, the common cradle of mankind lies in East or South Africa and can be dated back to roughly 100,000 years ago (see, e.g., Henn et al. (2012) for an overview). Starting from East Africa, a small sample of hunter-gatherers exited the African continent around 50,000-60,000 years ago and thereby started what is now also referred to as the "great human expansion". This expansion continued throughout Europe, Asia, Oceania, and the Americas, so that mankind eventually came to settle on all continents.

[1]The cross-country heterogeneity in patience and negative reciprocity is similar to, if not larger than, the heterogeneity in risk aversion and the prosocial traits, so that the weaker effects for the former preferences are not driven by a lack of variation.

[2]While we partly work with genetic distance, our objective is different from twin studies on economic preferences (Cesarini et al., 2009, 2012). Whereas these papers document that genes explain part of the variation in preferences, we do not aim at separating genetic from experience-based mechanisms, partly because the long-run scope of our approach in combination with recent evidence for gene-culture coevolution render such a "nature vs. nurture" endeavor misguided (Manuck and McCaffery, 2014; Henrich, 2015). Twin studies alone also cannot explain our results because the intergenerational transmission rates they imply are much too small to play a role for our long-run analysis.

A noteworthy feature of this very long-run process is that it occurred through a large number of discrete steps, each of which consisted of a sub-sample of the original population breaking apart and leaving the previous location to move on and found new settlements elsewhere. The main hypothesis underlying this paper is that the pattern of successive breakups and the resulting distribution of temporal distances across populations affected the distribution of economic preferences we observe around the globe today. After splitting apart, these sub-populations often settled geographically distant from each other, i.e., lived in separation. There are at least two channels through which the length of separation of two groups might have had an impact on between-group differences in preferences.[3]

First, if two populations have spent a long time apart from each other, they were subject to different historical experiences. Recent work highlights that economic preferences are malleable by idiosyncratic experiences or, more generally, by the composition of people's environment (Voors et al., 2012; Rao, 2018; Kosse et al., 2016; Callen et al., 2014; Alan and Ertac, Forthcoming). Thus, the differential historical experiences which have accumulated over thousands of years of separation might have given rise to different preferences as of today.

Second, whenever two populations spend time apart from each other, they develop different population-level genetic pools due to random genetic drift or location-specific selection pressures. Given that attitudes like risk aversion, trust, and altruism are transmitted across generations and that part of this transmission is genetic in nature (Cesarini et al., 2009; Dohmen et al., 2012), the different genetic endowments induced by long periods of separation could also generate differences in preferences.

We now formally illustrate how both of these channels (historical experiences and genetic pools) yield the prediction that longer separation implies larger absolute differences in preferences. For this purpose, we conceptualize both idiosyncratic 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 assump-

[3]It is conceivable that differences in preferences are correlated with temporal distance proxies because of the structure of the population breakups *as such*, rather than the temporal distances that were caused by the population breakups. Section 7 provides a discussion of this issue.

tions 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 $t = 0, 1, \ldots, T$, 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 $t = 0, 1, \ldots$ let $\mathscr{P}_t$ be a partition of $\{1, \ldots, N\}$, that is, $\mathscr{P}_t$ is a collection of disjoint nonempty sets whose union is $\{1, \ldots, N\}$. The elements of $\mathscr{P}_t$ represent the different populations at time t. For each $t \geq 0$ and $i \in \{1, \ldots, N\}$ let $P_t(i)$ be the unique $A \in \mathscr{P}_t$ 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 $A \in \mathscr{P}_t$ 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 $i, j \in \{1, \ldots, N\}$. 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 $i, j \in \{1, \ldots, N\}$ let $s_{ij} = |\{t \in \{1, \ldots, T\} : P_t(i) \neq P_t(j)\}|$. 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 nontrivial) 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, $A \in \mathscr{P}_t$, $t = 1, \ldots, T$, *are i.i.d. nondegenerate integrable random variables. Let* $i, j, k, l \in \{1, \ldots, N\}$. *Then*

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

The proof is in Appendix B.[4] 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.*

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.*

[4]We are deeply indebted to Lorens Imhof for proposing the proof to us.

3 Data

3.1 Risk, Time, and Social Preferences Across Countries

The data on risk, time, and social preferences are part of the Global Preference Survey (GPS), which constitutes a unique dataset on economic preferences from representative population samples around the globe. In a wide range of countries, the Gallup World Poll regularly surveys representative population samples about social and economic issues. In 76 countries, we included as part of the regular 2012 questionnaire a set of survey items which were explicitly designed to measure a respondent's preferences (see Falk et al., forthcoming, for a detailed description of the dataset).

Four noteworthy features characterize these data. First, the preference measures have been elicited in a comparable way using a standardized protocol across countries. Second, contrary to small- or medium-scale experimental work, we use preference measures of representative population samples in each country. This allows for inference on between-country differences in preferences, in contrast to existing cross-country comparisons of convenience (student) samples. The median sample size was 1,000 participants per country; in total, we collected preference measures for more than 80,000 participants worldwide. Respondents were selected through probability sampling and interviewed face-to-face or via telephone by professional interviewers. Third, the dataset also reflects geographical representativeness. The sample of 76 countries is not restricted to Western industrialized nations, but covers all continents and various development levels. Specifically, our sample includes 15 countries from the Americas, 24 from Europe, 22 from Asia and Pacific, as well as 14 nations in Africa, 11 of which are Sub-Saharan. The set of countries contained in the data covers about 90% of both the world population and global income. Fourth, the preference measures are based on experimentally validated survey items for eliciting preferences. In order to ensure behavioral relevance, the underlying survey items were designed, tested, and selected through an ex-ante experimental validation procedure (Falk et al., 2015). In this validation step, out of a large set of preference-related survey questions, those items were selected which jointly performed best in explaining observed behavior in standard financially incentivized experimental tasks to elicit preference parameters. In order to make these items cross-culturally applicable, (i) all items were translated back and forth by professionals, (ii) monetary values used in the survey were adjusted along the median household income for each country, and (iii) pretests were conducted in 21 countries of various cultural heritage to ensure comparability. The preference measures are derived as follows (see Appendix A and Falk et al. (forthcoming) for details):[5]

[5]The description of the survey items closely follows the one in Falk et al. (forthcoming).

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? Please indicate your answer on a scale from 0 to 10, where a 0 means "not willing to take risks at all" and a 10 means "very willing to take risks". You can also use the values in between to indicate where you fall on the scale."*

The quantitative measure is derived from a series of five interdependent hypothetical binary lottery choices, a format commonly referred to as the "staircase procedure". 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 being offered in the next question, and conversely. For instance, in Germany, the fixed upside of the lottery x was €300, and in the first question, the safe payment was €160. In case the respondent chose the lottery (the safe payment), the safe payment increased (decreased) to €240 (80) in the second question. In essence, by adjusting the safe payment according to previous choices, the questions "zoom in" around the respondent's certainty equivalent and make efficient use of limited and costly survey time. This procedure yields one of 32 ordered outcomes. The self-assessment and the outcome of the quantitative lottery staircase 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 between immediate and delayed financial rewards. In each of the five questions, participants had to decide between receiving a payment today or larger payments in 12 months. Conceptually 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 would be 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 (worth between €5 and €30 in €5 intervals) 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 these variables into a prosociality score that consists of the unweighted average of the three variables.[6]

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*.

Discussion: Heterogeneity, Stability, and Behavioral Relevance. As discussed in Falk et al. (forthcoming), the preference measures are constructed by linearly combining responses to the survey items using weights that are derived from the experimental validation procedure (Falk et al., 2015). See Appendix A.7.2 for details. All preference measures are then standardized to have mean zero and standard deviation of one. Falk et al. (forthcoming) show that all preferences exhibit a large amount of variation across countries. For example, calculating t-tests of all possible pairwise country comparisons reveals that about 80% of all country differences are statistically significant at the 1% level, for each preference.

[6]The country-level correlations between the three measures range between 0.27 and 0.71, see Falk et al. (forthcoming).

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 and standardize these variables again. 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 (multi-dimensional) preference dissimilarity, and hence as a proxy for cultural differences in contexts involving economic preference parameters.

Our objective of explaining preference differences through historical events implicitly assumes that preferences exhibit some degree of stability over time. While our data do not have a panel dimension, we can indirectly gauge population-level stability by comparing the preferences of young and old people. To this end, we compute the average preference among the young and old (split at age 40) and then correlate preferences of these two groups. If population-level preferences were very unstable, the correlation between young and old should be very low. However, in this exercise, the average correlation coefficient across preferences is $\rho = 0.91$, suggesting that preferences exhibit considerable population-level stability.

Finally, understanding the global variation in preferences is only meaningful to the extent that our measures capture behaviorally relevant heterogeneity. In this respect, the results in Falk et al. (forthcoming) provide encouraging evidence because preferences are correlated with those behaviors one would expect: For example, patience correlates with educational attainment and savings, risk taking with self-employment, and the social preferences with various social behaviors including donating, volunterring, and helping others. These within-country correlations are similar across countries.

3.2 Proxies for Ancient Migration Patterns

We use two different but conceptually linked classes of variables to proxy for the length of time since two populations split apart, i.e., genetic distance and linguistic distance.

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 ("noise") 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. In fact, akin to a molecular clock, population geneticists have made use of this observation by constructing mathematical models to compute the timing of separation between groups. This makes clear that, at its very core, genetic distance constitutes not only a measure of genealogical relatedness, but also of *temporal distance*.

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 et al. (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 et al. (1994) to countries.[7] 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.The key advantage of the genetic distance data relative to predicted measures of length of separation (see below) is that the measurement and imputation apply to *contemporary* populations. Thus, for example, the effects of smaller-scale migratory movements after the human exodus from Africa on the temporal distance between populations are by construction incorporated in these measures.

Recently, Spolaore and Wacziarg (2017) introduced a new dataset of cross-country F_{ST} genetic distances that is based on the work by Pemberton et al. (2013). While the data from Cavalli-Sforza et al. (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.

[7]To this end, the authors used ethnic composition data from Fearon (2003): the data by Cavalli-Sforza et al. (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 et al. (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.

Linguistic Distance. Population geneticists and linguists have long noted the close correspondence between genetic distance and linguistic "trees", intuitively because population break-ups do not only produce diverging gene pools, but also differential languages. Hence, we employ the degree to which two countries' languages differ from each other as an additional proxy for the timing of separation. The construction of linguistic distances 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.[8] For each country pair, we calculate the weighted linguistic distance according to the population shares speaking a particular language in the respective countries today.[9]

As a second and complementary measure of linguistic distance, we exploit 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 et al., 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 constucted by counting the number of phonetic edits needed to rewrite each word from one language spelling to another. That is, for each of the fourty words and each language pair, Wichmann et al. (2016) compute the number of phonetic edits, normalize this measure to account for word length, and average across words. 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

[8]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 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.

[9]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

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

the respective countries today.

Construction of Composite Measure of Temporal Distance. In sum, we have access to four proxies for temporal distance.[10] Given that these measures follow different methods of construction and are likely to be plagued by measurement error, we exploit the complementarity of the different data sources by constructing a composite index of temporal distance. This index is computed as unweighted average of the standardized values (z-scores) of the four distance variables. We standardize the temporal distance measure into a z-score to ease interpretation of regression coefficients.

4 Cross-Country Evidence

4.1 Regression Methodology

Since temporal distance is a bilateral variable, our analysis necessitates the use of a *dyadic* regression framework, which takes each possible pair of countries as unit of observation. Accordingly, we match each of the 76 countries with every other country into a total of 2,850 country pairs and, for each trait, compute the absolute difference in (average) preferences between the two countries.[11] We then relate our temporal distance measure to this absolute difference in preferences between the respective populations. Our regression equation is hence given by:

$$|\text{pref}_i - \text{pref}_j| = \alpha + \beta \times \text{temporal distance}_{i,j} + \gamma_i \times d_i + \gamma_j \times d_j + \epsilon_{i,j}$$

where pref_i and pref_j represent some average preference in countries i and j, respectively, d_i and d_j country fixed effects, and $\epsilon_{i,j}$ a country pair specific disturbance term.

As is standard practice in dyadic analyses such as in gravity regressions of bilateral trade, every specification to be presented below will include country fixed effects d_i and d_j, i.e., a fixed effect for each of the two countries that appears in a country pair observation to take out any unobservables that are country-specific.[12] To illustrate, with fixed effects for both countries, the regressions do not relate, say, the raw difference in preferences between Sweden and Mexico to the respective raw temporal distance. Rather, the regression relates the difference in preferences between Sweden and Mexico *relative* to Sweden's and Mexico's average differences in preferences in all country pairs to their temporal distance, again relative to all other temporal distances involving these

[10] Appendix C reports raw correlations among these proxies.

[11] Since the analysis is not directional, each country pair is only used once, i.e., when country i is matched with country j, j cannot be matched with i.

[12] Also see the working paper version of Spolaore and Wacziarg (2009).

two countries. For instance, if Mexico had very large differences in preferences to all countries, then the fixed effects would ensure that these uniform large differences are treated as a Mexico-specific effect, rather than attribute them to the bilateral relationships between Mexico and other countries. Thus, country-specific factors are netted out of the analysis and the regression equation estimates the bilateral effect of interest.[13]

Regarding the noise term, because our empirical approach implies that each country will appear multiple times as part of the (in)dependent variable, we need to allow for clustering of the error terms at the country-level. We hence employ the two-way clustering strategy of Cameron et al. (2011), i.e., we cluster at the level of the first and of the second country of a given pair. This procedure allows for arbitrary correlations of the error terms within a group, i.e., within the group of country pairs which share the same first country or which share the same second country, respectively.

4.2 Baseline Results

Table 1 provides the results of OLS regressions of absolute differences in preferences on temporal distance. For each dependent variable, we report two specifications. In the first column, the dependent variable is the absolute difference in average preferences in a country pair. In the second column, the dependent variable is the absolute difference in average *residual* preferences. Here, before aggregating preferences at the country level, we partial out age, age squared, gender, log household income p/c, educational attainment fixed effects, and marital status fixed effects.[14] Thus, this measure of difference in residual preferences reflects differences that are independent of variation along this comprehensive set of observables.

Throughout the paper, all regression coefficients (except for those of binary variables) are expressed in terms of standardized betas, i.e., both the dependent and the independent variables are normalized into z-scores and the dependent variable is then multiplied with 100, so that the coefficient can be interpreted as the percent change of a standard deviation in the dependent variable in response to a one standard deviation increase in the independent variable.

Columns (1) and (2) document that the summary statistic of preference differences

[13]The empirical results suggest that such country fixed effects indeed go a long way in addressing omitted variable concerns. For instance, in the analyses to be presented below, for patience and negative reciprocity we sometimes observe statistically significant *negative* coefficients on temporal distance if country fixed effects are not included, which we find very hard to interpret. These results entirely disappear with country fixed effects.

[14]Specifically, to construct this measure, we follow the following procedure. First, run an OLS regression of a given preference on the set of covariates described above. Second, compute the residuals of the regression. Third, aggregate the residuals to the country level. Fourth, compute the absolute difference in residual preferences for a given country pair.

Table 1: Preferences and temporal distance across countries

	Dependent variable: Absolute difference in...									
	All preferences		Risk taking		Prosociality		Neg. reciprocity		Patience	
	Raw	Residual	Raw	Residual	Raw	Residual	Raw	Residual	Raw	Residual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Temporal distance	0.22^{***}	0.17^{***}	0.14^{**}	0.14^{**}	0.14^{***}	0.088^{***}	0.038^{*}	0.031	0.11^{*}	0.081^{*}
	(0.05)	(0.04)	(0.07)	(0.07)	(0.04)	(0.03)	(0.02)	(0.03)	(0.06)	(0.05)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2850	2850	2850	2850	2850	2850	2850	2850	2850	2850
R^2	0.48	0.46	0.62	0.65	0.48	0.52	0.47	0.44	0.52	0.39

Notes. OLS estimates, twoway-clustered standard errors (clustered at both countries in a pair) in parentheses. The unit of observation is a country pair. The absolute difference in residual preferences is computed after individual-level preferences are residualized from age, age squared, gender, log household income p/c, educational attainment fixed effects, and marital status fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(which consists of the sum of the absolute differences across preference dimensions) is strongly and significantly related to temporal distance. The associated t-statistic equals 5.0 and the point estimate suggests that a one standard deviation increase in temporal distance is associated with an increase of 22 percent of a standard deviation in differences in preferences.

Columns (3) through (10) break this pattern into the separate preferences. The results are strongest for risk aversion and prosociality. For negative reciprocity and patience, the point estimates are always positive, but rather small in magnitude and only marginally significantly correlated with temporal distance. In sum, differences in all preferences are increasing in the length of separation of the respective populations, albeit to different degrees.

4.3 Separate Temporal Distance Proxies

To document that the relationship between differences in preferences and temporal distances is not driven by a particular measure, we proceed by relating differences in preferences to each temporal distance proxy separately. Table 2 documents that all variables that we use to proxy for temporal distance are strongly related to preference differences. Moreover, the standardized beta coefficients are all in the same ballpark and suggest that an increase in one standard deviation of temporal distance is associated with an increase in differences in preferences by 16-26% of a standard deviation.

4.4 Further Robustness Checks

Sub-Samples. A potential concern is that temporal distance might simply pick up regional effects. To address this, Appendix D.1 presents a set of regressions in which we exclude each continent one-by-one. This does not affect the results.

In many of the countries furthest from East Africa, the majority of the population is not indigenous. Our analysis addressed this aspect by employing observed genetic and linguistic distance as inputs into the explanatory variable, which by construction pertain to contemporary populations. Still, to rule out that the mass migration post-1500 and its effect on temporal distances drives our results, Appendix D.1 presents the results of an additional robustness check in which we restrict the sample to countries in the Old World, i.e., we exclude Australia, the Americas, and the Caribbean. Reassuringly, the results are very similar to the baseline results.

Culture-Dependent Interpretations. In most cases, our preference measures are composed of a combination of qualitative and quantitative survey items. While these items

Table 2: Robustness: Separate temporal distance proxies

	Dependent variable:			
	Absolute difference in all preferences			
	(1)	(2)	(3)	(4)
Fst genetic distance (Cavalli-Sforza)	0.26***			
	(0.04)			
Fst genetic distance (Pemberton et al.)		0.22***		
		(0.05)		
Linguistic distance (tree)			0.16***	
			(0.04)	
Linguistic distance (ASJP)				0.16***
				(0.04)
Country FE	Yes	Yes	Yes	Yes
Observations	2701	2628	2850	2850
R^2	0.48	0.47	0.48	0.47

Notes. OLS estimates, twoway-clustered standard errors (clustered at both countries in a pair) in parentheses. The unit of observation is a country pair. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

have been tailored to be applicable in cross-cultural research and have been subject to an extensive pre-test in 22 countries of different cultural heritages (see Falk et al. (forthcoming) for details), it is conceivable that the interpretation of the qualitative scales differs systematically across countries. If such a difference in interpretation was associated with temporal distance, our results might merely reflect contemporary differences in answering subjective self-assessments. The quantitative measures, on the other hand, are context-free hypothetical decisions over (purchasing power adjusted) monetary stakes in precisely defined choice contexts. Arguably, these measures do not suffer from the potential confound of being interpreted in different ways across countries. We hence replicate our main analysis using only quantitative preference measures. Appendix D.2 shows that the corresponding results are similar to, if not stronger, than those reported above.

Our survey items were selected based on an experimental validation procedure with German experimental subjects. To ensure that our temporal distance measure does not spuriously pick up differences in interpretation related to linguistic differences from Germany, Appendix D.3 shows that controlling for the relative linguistic distance to Germany between countries in a pair does not affect the results.

4.5 Multiple Testing

Strictly speaking, our empirical analysis is subject to multiple testing concerns because we evaluate the null hypothesis "temporal distance does not affect preference differences" through estimations that feature four dependent variables. At the same time, such concerns are arguably greatly reduced by our procedure of collapsing all dependent variables into a summary statistic. In doing so, we have only one regression specification to evaluate, and here overall preference differences are strongly related to temporal distance (see columns (1) and (2) of Table 1). We further address concerns about multiple testing in Appendix D.4 by presenting p-values which are adjusted using the false discovery rate (FDR) procedure (Anderson, 2012; Cantoni et al., 2017). Again, these results support the picture developed in the main text.

4.6 Conditional Regressions

The argument made in this paper is that the relationship between temporal distance on the one hand and preferences on the other hand reflects the impact of ancient migration patterns and the resulting distribution of temporal distances across populations, rather than *contemporary* differences in idiosyncratic country characteristics. We hence proceed by investigating the robustness of the relationship between temporal distance and preferences through conditional regressions. Throughout this section, it will be important to keep in mind that when we "control" for, say, geographic differences between countries, we are likely to "over-control". After all, differences in geographic and climatic conditions might be one of the channels through which temporal distance generates differences in preferences, in particular given that differences in climatic or geographic conditions are to a large extent very persistent. In what follows, our augmented regression specification will be:

$$|pref_i - pref_j| = \alpha + \beta \times \text{temporal distance}_{i,j} + \gamma_i \times d_i + \gamma_j \times d_j + \eta \times g_{i,j} + \epsilon_{i,j}$$

where g_{ij} is a vector of bilateral measures between countries i and j (such as their geodesic distance or the absolute difference in per capita income). Details on the definitions and sources of all control variables can be found in Appendix H.

We start by considering the summary statistic of preference differences. Throughout this analysis, we restrict the sample to country-pairs for which we have access to all covariates. To check that our coefficient of interest does not spuriously pick up the effect of demographic differences or differential population characteristics, column (2) of Table 3 adds to the baseline specification the absolute differences in proportion of females, religious fractionalization, and the fraction of the population who are of European descent. This joint set of covariates reduces the point estimate of temporal distance by only about 5%, and the coefficient remains statistically significant.

A potential concern with our baseline specification is that it ignores differences in development and institutions across countries, in particular given that temporal distance has been shown to correlate with differences in national income (Spolaore and Wacziarg, 2009). Column (3) of Table 3 therefore introduces absolute differences in (log) GDP per capita, democracy, and a common legal origin dummy. The inclusion of this vector of controls has almost no effect on the coefficient of interest.

To ensure that effects stemming from variations in geography or climate are not attributed to temporal distance, we condition on an exhaustive set of corresponding control variables. Column (4) introduces four distance metrics as additional controls into this regression. Our first geographical control variable consists of the geodesic distance (measuring the shortest distance between any two points on earth) between the most populated cities of the countries in a given pair. Relatedly, we introduce a dummy equal to one if two countries are contiguous. Finally, we also condition on the "distance" between two countries along the two major geographical axes, i.e., the difference in the distance to the equator and the longitudinal (east-west) distance. Again, the introduction of these variables has virtually no effect on the coefficient of temporal distance.

Of particular interest is perhaps the difference in coefficients between temporal distance and log geodesic distance. While geodesic distance is significantly correlated with preference differences in regressions that do not include temporal distance ($p < 0.01$), it ceases to have explanatory power once temporal distance is accounted for. These results are indicative that the precise migration patterns of our ancestors, rather than simple shortest-distance calculations between contemporary populations, need to be taken into account to understand the cross-country variation in preferences.

Given that geographic distance as such does not seem to drive our results, we now control for more specific information about differences in the micro-geographic and climatic conditions between the countries in a pair. To this end, we make use of information on the agricultural productivity of land, different features of the terrain, and climatic factors, see column (5). Again, the coefficient of temporal distance remains unaffected. Thus, the estimates remain remarkably robust across the different specifications. This suggests that – in order for omitted variable bias to explain our results – unobservables would have to bias our results by much more than the very large and comprehensive set of covariates in our regressions (Altonji et al., 2005; Bellows and Miguel, 2009).

Table 4 repeats the conditional regressions for all preferences separately. For each preference, we present the specification that includes all controls from column (5) of

Table 3: Preferences and temporal distance: Robustness (1/2)

	Dependent variable: Absolute difference in all preferences				
	(1)	(2)	(3)	(4)	(5)
Temporal distance	0.22***	0.21***	0.19***	0.19***	0.19***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Δ Proportion female		0.058	0.072	0.071	0.066
		(0.04)	(0.05)	(0.05)	(0.05)
Δ Religious fractionalization		0.0028	0.0040	0.00024	0.000055
		(0.03)	(0.03)	(0.03)	(0.03)
Δ % Of European descent		0.010	-0.045	-0.053*	-0.053*
		(0.03)	(0.03)	(0.03)	(0.03)
Δ Democracy index			-0.0017	0.0013	-0.0017
			(0.04)	(0.04)	(0.04)
Δ Log [GDP p/c PPP]			0.16***	0.15***	0.15***
			(0.05)	(0.05)	(0.05)
Log [Geodesic distance]				0.070	0.067
				(0.05)	(0.05)
1 for contiguity				0.012	0.013
				(0.02)	(0.02)
Δ Distance to equator				0.0028	0.0024
				(0.04)	(0.04)
Δ Longitude				-0.090*	-0.085*
				(0.05)	(0.04)
Δ Land suitability for agriculture					0.029
					(0.03)
Δ Mean elevation					0.0052
					(0.04)
Δ SD Elevation					-0.0100
					(0.03)
Δ Ave precipitation					-0.0045
					(0.04)
Δ Ave temperature					0.0038
					(0.04)
Δ Log [Area]					0.0088
					(0.03)
Country FE	Yes	Yes	Yes	Yes	Yes
Colonial relationship dummies	No	No	Yes	Yes	Yes
Observations	2556	2556	2556	2556	2556
R^2	0.47	0.47	0.49	0.49	0.49

Notes. OLS estimates, twoway-clustered standard errors (clustered at both countries in a pair) in parentheses. The unit of observation is a country pair. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Preferences and temporal distance: Robustness (2/2)

	Dependent variable: Absolute difference in...			
	Risk taking	Prosociality	Neg. recip.	Patience
	(1)	(2)	(3)	(4)
Temporal distance	0.14**	0.19***	-0.000090	0.0057
	(0.07)	(0.05)	(0.02)	(0.05)
Country FE	Yes	Yes	Yes	Yes
Population controls	Yes	Yes	Yes	Yes
Economic and institutional controls	Yes	Yes	Yes	Yes
Colonial relationship dummies	Yes	Yes	Yes	Yes
Distance controls	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes
Observations	2556	2556	2556	2556
R^2	0.62	0.48	0.49	0.58

Notes. OLS estimates, twoway-clustered standard errors (clustered at both countries in a pair) in parentheses. The unit of observation is a country pair. See Table 3 for a list of the covariates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. The relationship between risk preferences and prosociality on the one hand and temporal distance on the other hand is robust to this large and comprehensive vector of covariates, even quantitatively. At the same time, the effects on patience and negative reciprocity vanish once covariates are accounted for. Thus, consistent with the patterns reported above, it appears as if temporal distance has a stronger effect on risk preferences and prosocial traits than on patience and negative reciprocity.

5 Within-Country Evidence

5.1 Regression Methodology

Compared to between-country regressions, within-country analyses have the important advantage that they allow to hold constant many features of people's contemporary environments that are difficult to account for in cross-country analyses. This section makes further progress by considering variation in preferences and temporal distances within countries, across groups of migrants with different heritage. To this end, we use individual-level information about country of birth. In essence, these analyses will compare, say, the difference in preferences between French and Nigerians who currently live in the US with the difference between Italians and Japanese who also live in the US, or with the difference between Americans and Mexicans who live in the US. Thus, the unit of analysis is no longer a country pair, but rather a migrant-population-pair in a given country of residence.

Specifically, for 54 countries in our sample, we have information about the country of birth of our respondents. We compute the average level of a given preference at the country of residence times country of birth level, i.e., for each country of residence we compute the average preference for a given country of birth. In line with prior literature, we restrict the sample to migrant populations, although we have verified that including the native populations in a given country of residence does not affect the results. This procedure gives rise to 598 "populations". We match each population with each other population, but only keep those population pairs that share a common country of residence to be able to conduct a within-country analysis. Then, as before, we assign temporal distances to population pairs based on their countries of origin. Using this procedure, we end up with 6,232 population pairs from 144 countries of origin who currently live in 49 countries. Note that these statistics imply that the sample of populations is larger than in the cross-country analysis because we now observe migrants from countries which are not covered in the GPS. Our estimating equation is given by:

$$|\text{pref}_{i,z} - \text{pref}_{j,z}| = \alpha + \beta \times \text{temporal distance}_{i,j} + \gamma_i \times d_i + \gamma_j \times d_j + \gamma_z \times d_z + \epsilon_{i,j,z}$$

where $\text{pref}_{i,z}$ and $\text{pref}_{j,z}$ represent some average preference for people who currently reside in country z, yet were born in countries i and j, respectively. d_i and d_j are country of birth fixed effects. d_z are country of residence fixed effects, and $\epsilon_{i,j,z}$ a disturbance term. Thus, the regression equation is conceptually the same as in Section 4. As before, we employ a twoway-clustering strategy and cluster at the level of both countries of origin.

Working with sub-national groups comes at the cost that the number of respondents from any given country of birth in a given country of residence is sometimes very small, which implies that "population-level" preferences are measured with large error. To account for this, we restrict attention to population-pairs that consist of at least three respondents, i.e., in which one population consists of at least one and the other population of at least two respondents. This is arguably a very conservative procedure, yet still eliminates the most extreme forms of misattributing individual-level variation to population-level heterogeneity. This leaves us with 4,838 observations.[15]

[15]Table 12 in Appendix E reports a robustness check that includes all population-pair observations, i.e., even those that consist of only one respondent each. The results are slightly noisier, but overall very similar, in particular once variation in observables is accounted for.

Table 5: Preferences and temporal distance within country

	Dependent variable: Absolute difference in...									
	All preferences		Risk taking		Prosociality		Neg. reciprocity		Patience	
	Raw	Residual	Raw	Residual	Raw	Residual	Raw	Residual	Raw	Residual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Temporal distance	0.11^{***}	0.12^{***}	0.060^{**}	0.068^{**}	0.047^{*}	0.061^{**}	0.062^{*}	0.064^{**}	0.085^{***}	0.082^{***}
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Country of residence FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country of birth FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4838	4838	4827	4699	4802	4674	4775	4647	4803	4675
R^2	0.30	0.30	0.24	0.25	0.27	0.27	0.26	0.24	0.22	0.23

Notes. OLS estimates, twoway-clustered standard errors (clustered at both countries of origin) in parentheses. The unit of observation is a population pair, which is defined as two groups who currently reside in the same country, but were born in different countries. The absolute difference in residual preferences is computed after individual-level preferences are residualized from age, age squared, gender, log household income p/c, educational attainment fixed effects, and marital status fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Results

Table 5 presents the results. As in the cross-country case, we present two specifications for each preference. For each preference, in the first column the dependent variable is the raw absolute difference in preferences in a population pair. In the second column, the dependent variable is again the absolute difference in residual preferences. That is, as before, we make use of our individual-level covariates and residualize preferences from a set of observable characteristics. We then aggregate the residuals at the population level. This exercise is particularly important in the present within-country context because our analyses are based on relatively small samples. This implies the risk that differences in preferences across "populations" are confounded by individual characteristics, which introduces attenuation bias into our analysis.

The results document that, across all preferences, temporal distance is strongly and significantly related to differences in preferences. Thus, temporal distance is predictive of preference differences even among people who share the same contemporary country of residence. Interestingly, these results are even stronger than in the between-country case. Throughout, the coefficients are slightly larger in the regressions with residual preferences as dependent variable, which is consistent with variation in individual characteristics across small groups of migrants generating attenuation bias. We hence view the second "residual" specification as preferred one.

6 Evidence Across Linguistic Groups

6.1 Regression Methodology

Thus far, the analysis has taken as basic unit of observation a population as defined either by country of residence or by country of birth. In a final step, we exploit variation in temporal distance across linguistic groups. While countries are the natural and most common unit of observation in this line of research, it is evident that from the perspective of *temporal* distance, country borders are an artificial way to separate populations that have historically been united until very recently and still speak the same language, as, e.g., most Germans and Austrians. The cross-language analysis generalizes this idea. Because we do not know which linguistic group the GPS respondents actually belong to, we exploit information on respondents' interview language. In total, 82 different interview languages were used.[16] Appendix F provides an overview of the set of languages used in each country. We compute average preferences across

[16]The number of interview languages is too small to allow for a within-country analysis across linguistic groups. Given the dyadic regression logic, one needs at least three interview languages in a given country to generate within-country variation in linguistic distance.

individuals at the language level and then again generate a dyadic data set that consists of 3,321 pairs of languages. To illustrate, in this procedure, all respondents from Australia and Great Britain, and a subset of the respondents from Canada, Cameroon, Ghana, Kenya, Malawi, Nigeria, Rwanda, South Africa, the United States, Tanzania, Uganda, and Zimbabwe are lumped together in "English".

An important difference to the analyses above is that we cannot construct the full temporal distance index here because genetic distance data are not available at the language group level. Indeed, as is clear from the example of English above, there is large heterogeneity in genetic distance within a given language. Thus, our explanatory variable is the linguistic distance between two language groups. For the purposes of the analysis, we construct a composite measure of temporal distance, which consists of the average of the z-scores of the two linguistic distance variables described in Section 3, i.e., the language-tree based measure and the ASJP-based measure. The regression equation is given by:

$$|\text{pref}_i - \text{pref}_j| = \alpha + \beta \times \text{linguistic distance}_{i,j} + \gamma_i \times d_i + \gamma_j \times d_j + \epsilon_{i,j}$$

Now i and j refer to languages instead of countries, including the language fixed effects. There are two reasons to expect that the language-level analysis will suffer from measurement error and hence attenuation bias. First, Gallup's interview language need not necessarily correspond to the language that people actually speak in their daily lives. Second, as discussed in Section 3, in some cases people's language does not reflect their ancestral language (upon which meaningful measures of temporal distance can be based). While there is no way for us to correct for the first problem, we can address the second by excluding languages from the analysis for which the correspondence between language spoken and deep cultural heritage is particularly small, i.e., colonial languages. We hence present two types of regression specifications: one that uses the full sample of languages and one that excludes English, Spanish, and French.

6.2 Results

Table 6 presents the results. The summary statistic of preference differences is again significantly related to linguistic distance, both in the full sample of language-pairs and when we exclude the main colonial languages. Similar results again hold for risk taking and prosociality (columns (3)–(6)). The results are again weaker for patience and negative reciprocity. While the point estimates are consistently positive (compare columns (7)–(10)), they are not statistically significant. The slight difference in results between the cross-country and the cross-language analysis may be attributable to higher mea-

Table 6: Preferences and linguistic distance across linguistic groups

	Dependent variable: Absolute difference in...									
	All preferences		Risk taking		Prosociality		Neg. reciprocity		Patience	
Sample:	Full	Restricted	Full	Restricted	Full	Restricted	Full	Restricted	Full	Restricted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Linguistic distance	0.16***	0.17***	0.19***	0.19***	0.094**	0.098**	0.048	0.055	0.0040	0.0023
	(0.04)	(0.04)	(0.06)	(0.07)	(0.04)	(0.05)	(0.03)	(0.04)	(0.01)	(0.01)
Language FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3321	3081	3321	3081	3321	3081	3321	3081	3321	3081
R^2	0.55	0.53	0.56	0.56	0.44	0.42	0.52	0.51	0.60	0.61

Notes. OLS estimates, twoway-clustered standard errors (clustered at both languages in a pair) in parentheses. The unit of observation is a language pair. The restricted sample excludes all language pairs that include English, French, or Spanish. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

surement error and hence attenuation bias in the language-level analysis, e.g., because interview language does not perfectly map onto ethnolinguistic identity.

Table 13 in Appendix F.2 presents a robustness check on the baseline analysis in which we again residualize preferences from individual-level observables (age, age squared, gender, log household income p/c, educational attainment fixed effects, marital status fixed effects) before averaging them at the language level. The results are very similar to those established in the baseline analysis. Taken together, these results document that the relationship between preferences and temporal distance is not driven by (evolutionarily somewhat artifical) contemporary national borders.

7 Discussion

A growing body of empirical work highlights the importance of heterogeneity in risk, time, and social preferences for understanding a myriad of economic, social, and health behaviors. Indeed, not only are preferences correlated with individual-level behaviors, but also with country-level outcomes including comparative development, conflict, and institutional structures (Falk et al., forthcoming). Arguably, these correlations call for an understanding of the origins of the global variation in preferences. This paper has taken a step towards understanding these deep roots. Our main contribution is to establish that a significant fraction of the global variation in economic preferences has its historical origins in the structure and timing of very distant ancestral migration patterns, which highlights that if we aim to understand the ultimate roots of preference heterogeneity, we might have to consider events very far back in time. These results also bear an interesting relationship to work on cultural evolution (e.g., Boyd and Richerson, 1988; Henrich, 2015). In particular, our findings provide indirect evidence that preferences are indeed subject to evolutionary processes, as is assumed in models of cultural or genetic evolution. Our results also contibute to this debate in that they suggest that some preferences (risk aversion and prosociality) are to a larger extent driven by deep historical roots than others (patience and negative reciprocity).

Assessing the mechanisms underlying the relationship between temporal distance and preference differences is inherently difficult. First, the temporal distance variables capture variation that has accumulated over long time spans, which are characterized by poor data availability (and knowledge about living conditions in general). Second, a plethora of potential mechanisms could cause the effect of temporal distance on contemporary differences in preferences, including environmental influences, historical shocks, or genetic drift, to name but a few. In this respect, it also appears unlikely that a single event is responsible for explaining the connection between temporal distance and preferences, hence further complicating an analysis of mechanisms. Finally, it is reasonable to assume that the underlying processes and relative importance of channels differ across preferences.

With these caveats in mind, we investigate one particular potential mechanism, i.e., monotonic selective migration. As explained above, the great human expansion consisted of a succession of discrete migratory steps, in each of which subpopulations split apart from their parental colonies to found new settlements elsewhere. The model presented in Section 2 posits that preference differences between populations arose through *post-breakup* shocks, driven by, e.g., different experience. However, it is also conceivable that the *breakups per se* caused the patterns we observe if the new founder populations systematically differed from their parental colonies. This would be the case if, for example, only the least risk averse types tended to split away. In such a scenario, preferences would evolve *monotonically* along the migratory path out of East Africa, hence mechanically producing the correlation between temporal distance and preference differences. If true, this would still leave the main insight of the paper – that the structure and timing of population breakups in the very distant past have left a footprint in the contemporary global distribution of preferences – intact. However, the interpretation of this relationship would change slightly. Because we only observe preferences today, we cannot evaluate whether systematic population breakups actually took place. Still, what is relevant for our purposes is to investigate whether the results of such systematic breakups are still visible in the data today and hence drive our results.[17]

To investigate this issue, we regress the *level* of a given preference on (ancestry-adjusted) migratory distance from East Africa, i.e., Ethiopia. Table 14 in Appendix G provides an overview of the results and shows that our preference variables are not significantly correlated with migratory distance from East Africa. This pattern is indicative that – in line with our model – the relationship between temporal distance and preference differences is indeed driven by events *after* the various population breakups, rather than selective breakup patterns.

[17]Slightly more subtle, it is also possible that the correlation between temporal distance and preference differences is driven by a monotonic evolution of the *dispersion* of the preference pool, akin to the serial founder effect in population genetics: if the dispersion of the preference pool decreased monotonically along the migratory path, then differences in preferences between later founder populations would mechanically be smaller than those between earlier ones because the respective parental preference pool has lower variation to begin with. However, as shown in Appendix G.2, the relationship between preference dispersion and migratory distance from Ethiopia is very weak.

References

Alan, Sule and Seda Ertac, "Fostering Patience in the Classroom: Results from a Randomized Educational Intervention," *Journal of Political Economy*, Forthcoming.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," *Journal of Economic Growth*, 2003, *8* (2), 155–194.

__ **, Paola Giuliano, and Nathan Nunn**, "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 2013, *128* (2), 469–530.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2012, *103*, 1481–1495.

Ashraf, Quamrul and Oded Galor, "The Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 2013, *103* (1), 1–46.

__ **,** __ **, and Marc P. Klemp**, "The Out of Africa Hypothesis of Comparative Development Reflected by Nighttime Light Intensity," *Working Paper*, 2014.

Becker, Anke, "Herding and Male Dominance," *Working Paper*, 2017.

Bellows, John and Edward Miguel, "War and Local Collective Action in Sierra Leone," *Journal of Public Economics*, 2009, *93* (11), 1144–1157.

Bisin, Alberto and Thierry Verdier, "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 2001, *97* (2), 298–319.

Borghans, Lex and Bart HH Golsteyn, "Time Discounting and the Body Mass Index: Evidence from the Netherlands," *Economics & Human Biology*, 2006, *4* (1), 39–61.

Boyd, Robert and Peter J. Richerson, *Culture and the Evolutionary Process*, University of Chicago Press, 1988.

Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger, "Violence and Risk Preference: Experimental Evidence from Afghanistan," *American Economic Review*, 2014, *104* (1), 123–148.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Robust Inference with Multiway Clustering," *Journal of Business & Economic Statistics*, 2011, *29* (2), 238–249.

Cantoni, Davide, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Curriculum and Ideology," *Journal of Political Economy*, 2017, *125*, 338–392.

Cappelen, Alexander W., John A. List, Anya Samek, and Bertil Tungodden, "The Effect of Early Education on Social Preferences," *Working Paper*, 2016.

Cavalli-Sforza, Luigi, Paolo Menozzi, and Alberto Piazza, *The History and Geography of Human Genes*, Princeton University Press, 1994.

Cesarini, David, Christopher T. Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace, "Genetic Variation in Preferences for Giving and Risk Taking," *Quarterly Journal of Economics*, 2009, *124* (2), 809–842.

_ , **Magnus Johannesson, Patrik K.E. Magnusson, and Björn Wallace**, "The Behavioral Genetics of Behavioral Anomalies," *Management Science*, 2012, *58* (1), 21–34.

Chen, M. Keith, "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets," *American Economic Review*, 2013, *103* (2), 690–731.

Cohn, Alain and Michel A. Maréchal, "Laboratory Measure of Cheating Predicts Misbehavior at School," *Economic Journal*, forthcoming.

Desmet, Klaus, Michel Le Breton, Ignacio Ortuno-Ortin, and Shlomo Weber, "The Stability and Breakup of Nations: A Quantitative Analysis," *Journal of Economic Growth*, 2011, *16*, 183–213.

Doepke, Matthias and Fabrizio Zilibotti, "Culture, Entrepreneurship, and Growth," in "Handbook of Economic Growth" 2013, chapter 2.

_ **and** _ , "Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission," *Econometrica*, 2017, *85*, 1331–1371.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, "The Intergenerational Transmission of Risk and Trust Attitudes," *Review of Economic Studies*, 2012, *79* (2), 645–677.

Durante, Ruben, "Risk, Cooperation and the Economic Origins of Social Trust: An Empirical Investigation," *Working Paper*, 2009.

Enke, Benjamin, "Kinship Systems, Cooperation and the Evolution of Culture," *Working Paper*, 2017.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic Preferences," *Quarterly Journal of Economics*, forthcoming.

_ , _ , _ , David Huffman, and Uwe Sunde, "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences," *Working Paper*, 2015.

Fearon, James D., "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 2003, *8* (2), 195–222.

Fernández, Raquel and Alessandra Fogli, "Fertility: The Role of Culture and Family Experience," *Journal of the European Economic Association*, 2006, *4* (2-3), 552–561.

Gächter, Simon and Jonathan F. Schulz, "Intrinsic Honesty and the Prevalence of Rule Violations Across Societies," *Nature*, 2016, *531* (7595), 496–499.

Galor, Oded and Ömer Özak, "The Agricultural Origins of Time Preference," *American Economic Review*, 2016, *106* (10), 3064–3103.

Giuliano, Paola, "Living Arrangements in Western Europe: Does Cultural Origin Matter?," *Journal of the European Economic Association*, 2007, *5* (5), 927–952.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Cultural Biases in Economic Exchange?," *Quarterly Journal of Economics*, 2009, *124* (3), 1095–1131.

Haushofer, Johannes, Daniel Schunk, and Ernst Fehr, "Negative Income Shocks Increase Discount Rates," *Working Paper*, 2013.

Henn, Brenna M., Luigi L. Cavalli-Sforza, and Marcus W. Feldman, "The Great Human Expansion," *Proceedings of the National Academy of Sciences*, 2012, *109* (44), 17758–17764.

Henrich, Joseph, *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*, Princeton University Press, 2015.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter, "Antisocial Punishment Across Societies," *Science*, 2008, *319* (5868), 1362–1367.

Holman, Eric W., Cecil H. Brown, Søren Wichmann, André Müller, Viveka Velupillai, Harald Hammarström, Sebastian Sauppe, Hagen Jung, Dik Bakker, Pamela Brown et al., "Automated Dating of the World's Language Families Based on Lexical Similarity," *Current Anthropology*, 2011, *52* (6), 841–875.

Kosfeld, Michael and Devesh Rustagi, "Leader Punishment and Cooperation in Groups: Experimental Field Evidence from Commons Management in Ethiopia," *American Economic Review*, 2015, *105* (2), 747–783.

Kosse, Fabian, Thomas Deckers, Hannah Schildberg-Hörisch, and Armin Falk, "The Formation of Human Prosociality: Causal Evidence on the Role of Sociall Environment," *Working Paper*, 2016.

Manuck, Stephen B. and Jeanne M. McCaffery, "Gene-Environment Interaction," *Annual Review of Psychology*, 2014, *65*, 41–70.

Michalopoulosa, Stelios, "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 2012, *102* (4), 1508–1539.

Nordhaus, William D., "Geography and Macroeconomics: New Data and New Findings," *Proceedings of the National Academy of Sciences of the United States of America*, 2006, *103* (10), 3510–3517.

Olsson, Ola and Christopher Paik, "Long-Run Cultural Divergence: Evidence from the Neolithic Revolution," *Journal of Development Economics*, 2016, *122*, 197–213.

Özak, Ömer, "Distance to the Technological Frontier and Economic Development," *Working Paper*, 2016.

Pemberton, Trevor J., Michael DeGiorgio, and Noah A. Rosenberg, "Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation," *G3: Genes| Genomes| Genetics*, 2013, *3*, 891–907.

Putterman, Louis and David N. Weil, "Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 2010, *125* (4), 1627–1682.

Rao, Gautam, "Familiarity Does Not Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools," *Working Paper*, 2018.

Spolaore, Enrico and Romain Wacziarg, "The Diffusion of Development," *Quarterly Journal of Economics*, 2009, *124* (2), 469–529.

__ **and** __ , "Ancestry, Language and Culture," in Victor Ginsburgh and Shlomo Weber, eds., *The Palgrave Handbook of Economics and Language*, London: Palgrave Macmillan, 2016, chapter 6, pp. 174–211.

__ **and** __ , "Ancestry and Development: New Evidence," *Working Paper*, 2017.

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rützler, and Stefan T. Trautmann, "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior," *American Economic Review*, 2013, *103* (1), 510–531.

Voigtländer, Nico and Hans-Joachim Voth, "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany," *Quarterly Journal of Economics*, 2012, *127* (3), 1339–1392.

Voors, Maarten J., Eleonora E.M. Nillesen, Philip Verwimp, Erwin H. Bulte, Robert Lensink, and Daan P. Van Soest, "Violent Conflict and Behavior: A Field Experiment in Burundi," *American Economic Review*, 2012, *102* (2), 941–964.

Wichmann, Søren, Eric W. Holman, and Cecil H. Brown, "The ASJP Database (version 17)," *URL: http://email. eva. mpg. de/˜ wichmann/ASJPHomePage. htm*, 2016, *3*.

ONLINE APPENDIX

A Details on Global Preference Survey

This Appendix is taken from Falk et al. (forthcoming) for the convenience of the reader.

A.1 Overview

The cross-country dataset measuring risk aversion, patience, positive and negative reciprocity, altruism, and trust, was collected through the professional infrastructure of the Gallup World Poll 2012. The data collection process consisted of three steps. First, an experimental validation procedure was conducted to select the survey items. Second, there was a pre-test of the selected survey items in a variety of countries to ensure implementability in a culturally diverse sample. Third, the final data set was collected through the regular professional data collection efforts in the framework of the World Poll 2012.

A.2 Experimental Validation

To ensure the behavioral validity of the preference measures, all underlying survey items were selected through an experimental validation procedure (see Falk et al. (2015) for details). To this end, a sample of 409 German undergraduates completed standard state-of-the-art financially incentivized laboratory experiments designed to measure risk aversion, patience, positive and negative reciprocity, altruism, and trust. The same sample of subjects then completed a large battery of potential survey items. In a final step, for each preference, those survey items were selected which jointly performed best in explaining the behavior under real incentives observed in the choice experiments.

A.3 Pre-Test and Adjustment of Survey Items

Prior to including the preference module in the Gallup World Poll 2012, it was tested in the field as part of the World Poll 2012 pre-test, which was conducted at the end of 2011 in 22 countries. The main goal of the pre-test was to receive feedback on each item from various cultural backgrounds in order to assess potential difficulties in understanding and differences in the respondents' interpretation of items. Based on respondents' feedback and suggestions, minor modifications were made to several items before running the survey as part of the World Poll 2012.

The pre-test was run in 10 countries in central Asia (Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Russia, Tajikistan, Turkmenistan, Uzbekistan) 2 countries in South-East Asia (Bangladesh and Cambodia), 5 countries in Southern and Eastern Europe (Croatia, Hungary, Poland, Romania, Turkey), 4 countries in the Middle East and North Africa (Algeria, Jordan, Lebanon, and Saudi-Arabia), and 1 country in Eastern Africa (Kenya). In each country, the sample size was 10 to 15 people. Overall, more than 220 interviews were conducted. In most countries, the sample was mixed in terms of gender, age, educational background, and area of residence (urban / rural).

Participants in the pre-test were asked to state any difficulties in understanding the items and to rephrase the meaning of items in their own words. If they encountered difficulties in understanding or interpreting items, respondents were asked to make suggestions on how to modify the wording of the item in order to attain the desired meaning.

Overall, the understanding of both the qualitative items and the quantitative items was satisfactory. In particular, no interviewer received any complaints regarding difficulties in assessing the quantitative questions or understanding the meaning of the probability used in the hypothetical risky choice items. When asked about rephrasing the qualitative items in their own words, most participants seemed to have understood the items in exactly the way that was intended. Nevertheless, some (sub-groups of) participants suggested adjustments to the wording of some items. This resulted in minor changes to four items, relative to the "original" experimentally validated items:

1. The use of the term "lottery" in hypothetical risky choices was troubling to some Muslim participants. As a consequence, we dropped the term "lottery" and replaced it with "draw".

2. The term "charity" caused confusion in Eastern Europe and Central Asia, so it was replaced it with "good cause".

3. Some respondents asked for a clarification of the question asking about one's willingness to punish unfair behavior. This feedback lead to splitting the question into two separate items, one item asking for one's willingness to punish unfair behavior towards others, and another asking for one's willingness to punish unfair behavior towards oneself.

4. When asked about hypothetical choices between monetary amounts today versus larger amounts one year later, some participants, especially in countries with current or relatively recent phases of volatile and high inflation rates, stated that their answer would depend on the rate of inflation, or said that they would always

take the immediate payment due to uncertainty with respect to future inflation. Therefore, we decided to add the following phrase to each question involving hypothetical choices between immediate and future monetary amounts: "Please assume there is no inflation, i.e., future prices are the same as today's prices."

A.4 Selection of Countries

The goal when selecting countries was to ensure representative coverage of the global population. Thus, countries from each continent and each region within continents were chosen. Another goal was to maximize variation with respect to observables, such as GDP per capita, language, historical and political characteristics, or geographical location and climatic conditions. Accordingly, the selection process favored non-neighboring and culturally dissimilar countries. This procedure resulted in the following sample of 76 countries:

East Asia and Pacific: Australia, Cambodia, China, Indonesia, Japan, Philippines, South Korea, Thailand, Vietnam

Europe and Central Asia: Austria, Bosnia and Herzegovina, Croatia, Czech Republic, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Italy, Kazakhstan, Lithuania, Moldova, Netherlands, Poland, Portugal, Romania, Russia, Serbia, Spain, Sweden, Switzerland, Turkey, Ukraine, United Kingdom

Latin America and Caribbean: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Guatemala, Haiti, Mexico, Nicaragua, Peru, Suriname, Venezuela

Middle East and North Africa: Algeria, Egypt, Iran, Iraq, Israel, Jordan, Morocco, Saudi Arabia, United Arab Emirates

North America: United States, Canada

South Asia: Afghanistan, Bangladesh, India, Pakistan, Sri Lanka

Sub-Saharan Africa: Botswana, Cameroon, Ghana, Kenya, Malawi, Nigeria, Rwanda, South Africa, Tanzania, Uganda, Zimbabwe

A.5 Sampling and Survey Implementation

A.5.1 Background

Since 2005, the international polling company Gallup has conducted an annual World Poll, in which it surveys representative population samples in almost every country around the world on, e.g., economic, social, political, and environmental issues. The collection of our preference data was embedded into the regular World Poll 2012 and hence made use of the pre-existing polling infrastructure of one of the largest professional polling institutes in the world.[18]

Selecting Primary Sampling Units

In countries in which face-to-face interviews are conducted, the first stage of sampling is the identification of primary sampling units (PSUs), consisting of clusters of households. PSUs are stratified by population size and / or geography and clustering is achieved through one or more stages of sampling. Where population information is available, sample selection is based on probabilities proportional to population size. If population information is not available, Gallup uses simple random sampling.

In countries in which telephone interviews are conducted, Gallup uses a random-digit-dialing method or a nationally representative list of phone numbers. In countries with high mobile phone penetration, Gallup uses a dual sampling frame.

Selecting Households and Respondents

Gallup uses random route procedures to select sampled households. Unless an outright refusal to participate occurs, interviewers make up to three attempts to survey the sampled household. To increase the probability of contact and completion, interviewers make attempts at different times of the day, and when possible, on different days. If the interviewer cannot obtain an interview at the initially sampled household, he or she uses a simple substitution method.

In face-to-face and telephone methodologies, random respondent selection is achieved by using either the latest birthday or Kish grid methods.[19] In a few Middle East and Asian countries, gender-matched interviewing is required, and probability sampling with quotas is implemented during the final stage of selection. Gallup implements quality control procedures to validate the selection of correct samples and that the correct person is randomly selected in each household.

[18] See http://www.gallup.com/strategicconsulting/156923/worldwide-research-methodology.aspx

[19] The latest birthday method means that the person living in the household whose birthday among all persons in the household was the most recent (and who is older than 15) is selected for interviewing. With the Kish grid method, the interviewer selects the participants within a household by using a table of random numbers. The interviewer will determine which random number to use by looking at, e.g., how many households he or she has contacted so far (e.g., household no. 8) and how many people live in the household (e.g., 3 people, aged 17, 34, and 36). For instance, if the corresponding number in the table is 7, he or she will interview the person aged 17.

Sampling Weights
Ex post, data weighting is used to ensure a nationally representative sample for each country and is intended to be used for calculations within a country. These sampling weights are provided by Gallup. First, base sampling weights are constructed to account for geographic oversamples, household size, and other selection probabilities. Second, post-stratification weights are constructed. Population statistics are used to weight the data by gender, age, and, where reliable data are available, education or socioeconomic status.

A.5.2 Translation of Items

The items of the preference module were translated into the major languages of each target country. The translation process involved three steps. As a first step, a translator suggested an English, Spanish or French version of a German item, depending on the region. A second translator, being proficient in both the target language and in English, French, or Spanish, then translated the item into the target language. Finally, a third translator would review the item in the target language and translate it back into the original language. If differences between the original item and the back-translated item occurred, the process was adjusted and repeated until all translators agreed on a final version.

A.5.3 Adjustment of Monetary Amounts in Quantitative Items

All items involving hypothetical monetary amounts were adjusted for each country in terms of their real value. Monetary amounts were calculated to represent the same share of a country's median income in local currency as the share of the amount in Euro of the German median income since the validation study had been conducted in Germany. Monetary amounts used in the validation study with the German sample were "round" numbers to facilitate easy calculations (e.g., the expected return of a lottery with equal chances of winning and losing) and to allow for easy comparisons (e.g., 100 Euro today versus 107.50 in 12 months). To proceed in a similar way in all countries, monetary amounts were always rounded to the next "round" number. For example, in the quantitative items involving choices between a lottery and varying safe options, the value of the lottery was adjusted to a round number. The varying safe options were then adjusted proportionally as in the original version. While this necessarily resulted in some (very minor) variations in the real stake size between countries, it minimized cross-country differences in the understanding the quantitative items due to difficulties in assessing the involved monetary amounts.

A.6 Wording of Survey Items

In the following, "willingness to act" indicates the following introduction: *We now ask for your willingness to act in a certain way in four different areas. Please again indicate your answer on a scale from 0 to 10, where 0 means you are "completely unwilling to do so" and a 10 means you are "very willing to do so". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.*

Similarly, "self-assessments" indicate that the respective statement was preceded by the following introduction: *How well do the following statements describe you as a person? Please indicate your answer on a scale from 0 to 10. A 0 means "does not describe me at all" and a 10 means "describes me perfectly". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.*

A.6.1 Patience

1. (Sequence of five interdependent quantitative questions:) *Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose. Please assume there is no inflation, i.e, future prices are the same as today's prices. Please consider the following: Would you rather receive 100 Euro today or x Euro in 12 months?*

   The precise sequence of questions was given by the "tree" logic in Figure 1.

2. (Willingness to act:) *How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?*

A.6.2 Risk Taking

1. *(Similar to self-assessment:) Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.*

2. (Sequence of five interdependent quantitative questions:) *Please imagine the following situation. You can choose between a sure payment of a particular amount of money, or a draw, where you would have an equal chance of getting amount x or getting nothing. We will present to you five different situations. What would you prefer: a draw with a 50 percent chance of receiving amount x, and the same 50*

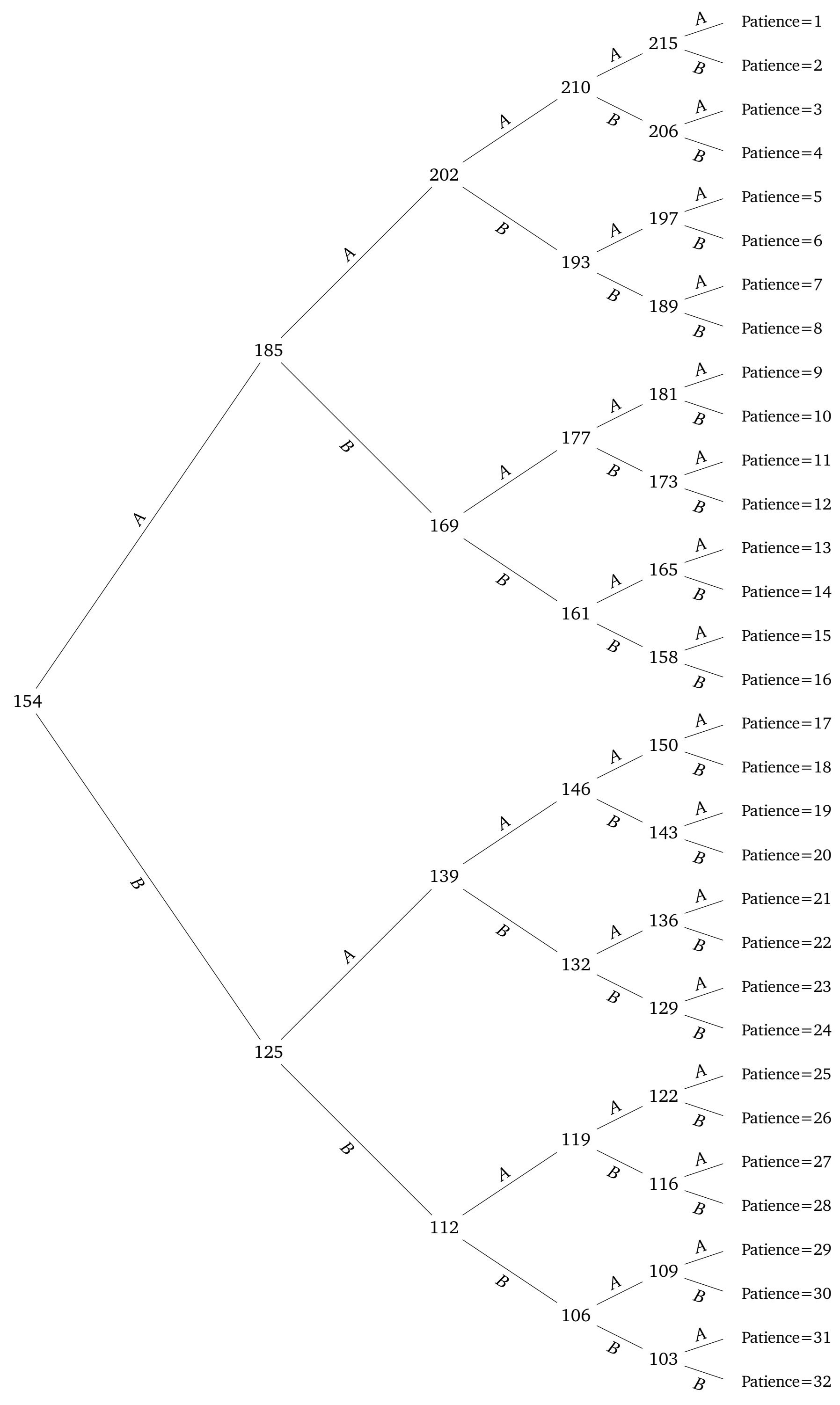

Figure 1: Tree for the staircase time task (numbers = payment in 12 months, A = choice of "100 euros today", B = choice of "x euros in 12 months". The staircase procedure worked as follows. First, each respondent was asked whether they would prefer to receive 100 euros today or 154 euros in 12 months from now (leftmost decision node). In case the respondent opted for the payment today ("A"), in the second question the payment in 12 months was adjusted upwards to 185 euros. If, on the other hand, the respondent chose the payment in 12 months, the corresponding payment was adjusted down to 125 euros. Working further through the tree follows the same logic.

percent chance of receiving nothing, or the amount of y as a sure payment? The precise sequence of questions was given by the "tree" logic in Figure 2.

A.6.3 Positive Reciprocity

1. (Self-assessment:) *When someone does me a favor I am willing to return it.*

2. *(Hypothetical situation:) Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. Helping you costs the stranger about 20 Euro in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 Euro, the most expensive one costs 30 Euro. Do you give one of the presents to the stranger as a "thank-you"-gift? If so, which present do you give to the stranger? No present / The present worth 5 / 10 / 15 / 20 / 25 / 30 Euro.*

A.6.4 Negative Reciprocity

1. (Self-assessment:) *If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so.*

2. (Willingness to act:) *How willing are you to punish someone who treats you unfairly, even if there may be costs for you?*

3. (Willingness to act:) *How willing are you to punish someone who treats others unfairly, even if there may be costs for you?*

A.6.5 Altruism

1. *(Hypothetical situation:) Imagine the following situation: Today you unexpectedly received 1,000 Euro. How much of this amount would you donate to a good cause? (Values between 0 and 1000 are allowed.)*

2. (Willingness to act:) *How willing are you to give to good causes without expecting anything in return?*

A.6.6 Trust

(Self-assessment:) *I assume that people have only the best intentions.*

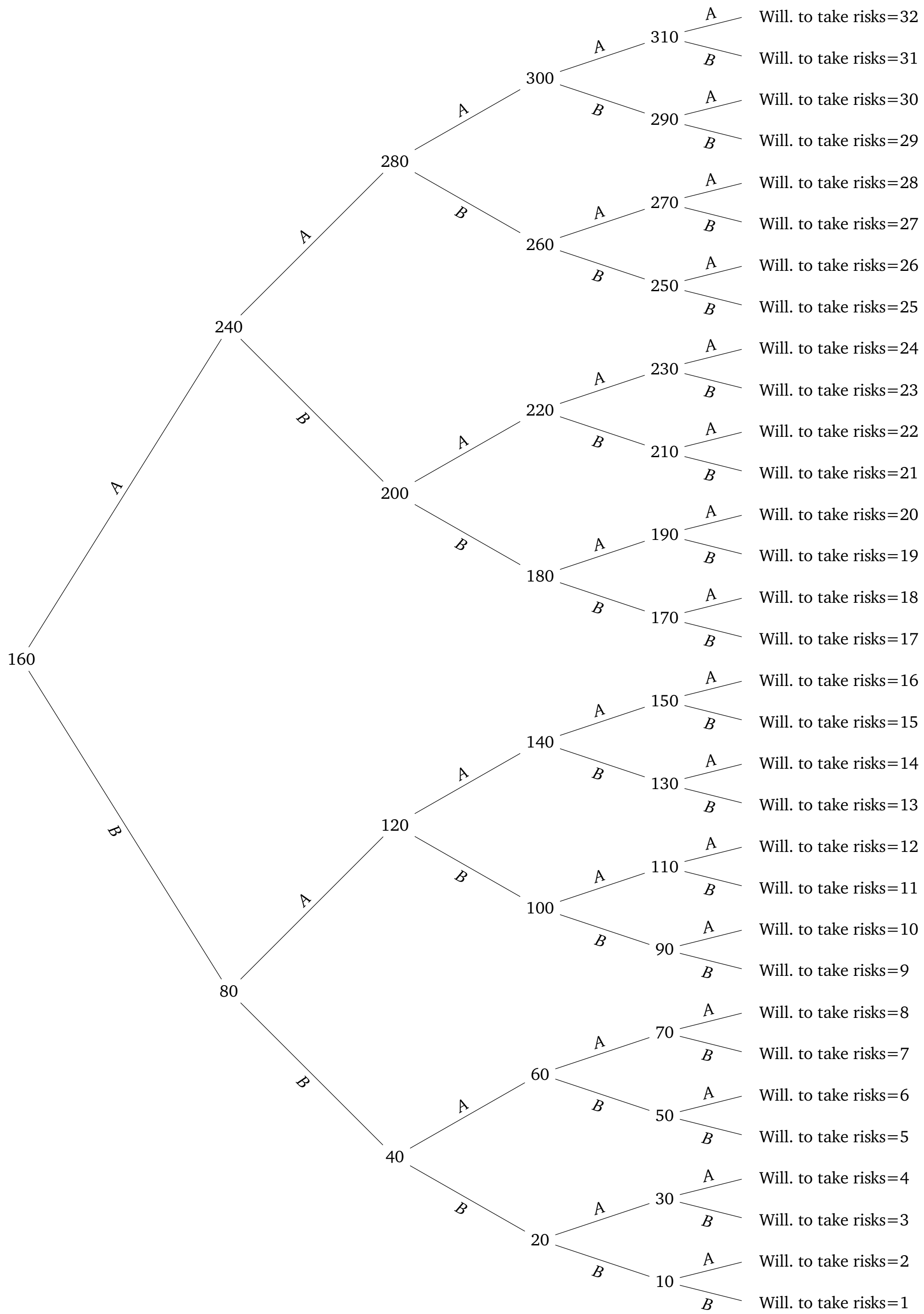

Figure 2: Tree for the staircase risk task (numbers = sure payment, A = choice of sure payment, B = choice of lottery). The staircase procedure worked as follows. First, each respondent was asked whether they would prefer to receive 160 euros for sure or whether they preferred a 50:50 chance of receiving 300 euros or nothing. In case the respondent opted for the safe choice ("B"), the safe amount of money being offered in the second question decreased to 80 euros. If, on the other hand, the respondent opted for the gamble ("A"), the safe amount was increased to 240 euros. Working further through the tree follows the same logic.

A.7 Computation of Preference Measures

A.7.1 Cleaning and Imputation of Missings

In order to efficiently use all available information in our data, missing survey items were imputed based on the following procedure:

- If one (or more) survey items for a given preference were missing, then the missing items were predicted using the responses to the available items. The procedure was as follows:
  - Suppose the preference was measured using two items, call them a and b. For those observations with missing information on a, the procedure was to predict its value based on the answer to b and its relationship to a, which was estimated by regressing b on a for the sub-sample of subjects who had nonmissing information on both, a and b (on the world sample).
  - For the unfolding-brackets time and risk items, the imputation procedure was similar, but made additional use of the informational content of the responses of participants who started but did not finish the sequence of the five questions. Again suppose that the preference is measured using two items and suppose that a (the staircase measure) is missing. If the respondent did not even start the staircase procedure, then imputation was done using the methodology described above. On the other hand, if the respondent answered between one and four of the staircase questions, a was predicted using a different procedure. Suppose the respondent answered four items such that his final staircase outcome would have to be either x or y. A probit was run of the "x vs. y" decision on b, and the corresponding coefficients were used to predict the decision for all missings (note that this constitutes a predicted probability). The expected staircase outcome was then obtained by applying the predicted probabilities to the respective staircase endpoints, i.e., in this case x and y. If the respondent answered three (or less) questions, the same procedure was applied, the only difference being that in this case the obtained predicted probabilities were applied to the expected values of the staircase outcome conditional on reaching the respective node. Put differently, the procedure outlined above was applied recursively by working backwards through the "tree" logic of the staircase procedure, resulting in an expected value for the outcome node.
  - If all survey items for a given preference were missing, then no imputation took place.

- Across the 12 survey items, between 0% and 8% of all responses had to be imputed.

A.7.2 Computation of Preference Indices at the Individual Level

For each of the traits (risk preferences, time preferences, positive reciprocity, negative reciprocity, altruism, and trust), an individual-level index was computed that aggregated responses across different survey items. Each of these indices was computed by (i) computing the z-scores of each survey item at the individual level and (ii) weighing these z-scores using the weights resulting from the experimental validation procedure of Falk et al. (2015). Formally, these weights are given by the coefficients of an OLS regression of observed behavior in the experimental validation study on responses to the respective survey items, such that the weights sum to one. In practice, for almost all preferences, the coefficients assign roughly equal weight to all corresponding survey items. The weights are given by:

$$
\begin{aligned}
\text{Patience} &= 0.7115185 \times \text{ Staircase patience } + 0.2884815 \times \text{ Will. to give up sth. today} \\
\text{Risk taking} &= 0.4729985 \times \text{ Staircase risk } + 0.5270015 \times \text{ Will. to take risks} \\
\text{Pos. reciprocity} &= 0.4847038 \times \text{ Will. to return favor } + 0.5152962 \times \text{ Size of gift} \\
\text{Neg. reciprocity} &= 0.6261938/2 \times \text{ Will. to punish if oneself treated unfairly} \\
&\quad + 0.6261938/2 \times \text{ Will. to punish if other treated unfairly} \\
&\quad + 0.3738062 \times \text{ Will. to take revenge} \\
\text{Altruism} &= 0.6350048 \times \text{ Will. to give to good causes } + 0.3649952 \times \text{ Hypoth. donation} \\
\text{Trust} &: \text{The survey included only one corresponding item.}
\end{aligned}
$$

A.7.3 Computation of Country Averages

In order to compute country-level averages, individual-level data were weighted with the sampling weights provided by Gallup, see above. These sampling weights ensure that our measures correctly represent the population at the country level.

B Proofs

Proof of Proposition 1. We have

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

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^t_A. 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,

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

and

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

The claimed equivalence will follow if we can show that

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

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,

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

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

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

$$E[|y+z||y] \geq |E[y+z|y]| = |E[y|y]| = |y|. \tag{2}$$

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 $E[|y+z|] > E[|y|]$. □

C Raw Correlations Among Temporal Distance Proxies

Table 7: Raw correlations among temporal distance proxies

	Fst dist.	Fst dist. (new)	Linguistic dist. (tree)	Linguistic dist. (ASJP)
Fst genetic dist. (Cavalli-Sforza)	1			
Fst genetic dist. (Pemberton et al.)	0.849	1		
Ling. dist. (tree)	0.443	0.370	1	
Ling. dist. (ASJP)	0.381	0.290	0.918	1

Pearson raw correlations. Fst dist. (new) refers to the Fst genetic distance measure based on Spolaore and Wacziarg (2017). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Additional Country-Pair-Level Regressions

D.1 Sub-Samples

Table 8: Robustness: Sub-samples

	Dependent variable: Abs. difference in all preferences				
	Sample excludes...				
	Europe & Central Asia	Sub-Saharan Africa & Middle East	South and East Asia & Pacific	Americas	New World
	(1)	(2)	(3)	(4)	(5)
Temporal distance	0.22***	0.19**	0.22***	0.27***	0.27***
	(0.05)	(0.08)	(0.05)	(0.05)	(0.06)
Country FE	Yes	Yes	Yes	Yes	Yes
Observations	1176	1540	1891	1830	1711
R^2	0.50	0.49	0.50	0.46	0.46

Notes. Country-level OLS estimates, twoway-clustered standard errors (clustered at both countries in a pair) in parentheses. The unit of observation is a country pair. In each column, the sample excludes a given world region. For example, in column (5), we exclude countries in the new world, , i.e., Australia, the Americas, and the Caribbean. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Using Quantitative Preference Measures Only

We exclude qualitative preference measures and run our baseline analysis using quantitative measures only which are available for risk taking, patience, altruism, and positive reciprocity. Table 9 shows that the results remain unchanged. In fact, they become much stronger in the case of patience.

D.3 Temporal Distance to Germany

Table 10 checks robustness against including relative linguistic distance to Germany.

D.4 Adjusting p-Values Using the FDR Procedure

This section reports p-values that are adjusted for multiple testing using the FDR procedure (see Anderson, 2012; Cantoni et al., 2017, for details). To assess the null hypothesis “temporal distance does not affect preference differences”, we group the regressions across dependent variables. Table 11 presents adjusted p-values for the same regressions as in Table 1. Note that adjusted p-values can be smaller than adjusted ones.

Table 9: Quantitative preference measures and temporal distance across countries

	Dependent variable: Absolute difference in...				
	All quant. measures	Staircase risk	Donation amount	Size of gift	Staircase patience
	(1)	(2)	(3)	(4)	(5)
Temporal distance	0.18^{***}	0.18^{**}	0.039^{*}	0.057	0.13^{**}
	(0.05)	(0.07)	(0.02)	(0.04)	(0.06)
Country FE	Yes	Yes	Yes	Yes	Yes
Observations	2850	2850	2850	2850	2850
R^2	0.53	0.53	0.70	0.51	0.52

Notes. Country-level OLS estimates, twoway-clustered standard errors (clustered at both countries in a pair) in parentheses. The unit of observation is a country pair. The dependent variables are the absolute differences in responses to the quantitative survey items. In column (1), the dependent variable is the average absolute difference of all quantitative items. In column (2), it is the outcome of staircase risk procedure, which measures risk aversion. In column (3), it is the donation amount that is part of the altruism variable. In column (4), the dependent variable is the size of a gift, which is an input into the positive reciprocity variable. In column (5), the dependent variable is the outcome of the staircase patience procedure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Robustness: Temporal distance to Germany

	Dependent variable: Absolute difference in...				
	All prefs.	Risk taking	Prosociality	Neg. reciprocity	Patience
	(1)	(2)	(3)	(4)	(5)
Temporal distance	0.20***	0.15**	0.13***	0.041	0.032
	(0.05)	(0.07)	(0.05)	(0.03)	(0.04)
Relative temporal distance to Germany	0.040	-0.011	0.011	-0.0050	0.12*
	(0.05)	(0.03)	(0.04)	(0.05)	(0.06)
Country FE	Yes	Yes	Yes	Yes	Yes
Observations	2850	2850	2850	2850	2850
R^2	0.49	0.62	0.49	0.47	0.53

Notes. OLS estimates, twoway-clustered standard errors (clustered at both countries in a pair) in parentheses. The unit of observation is a country pair. Relative temporal distance to Germany is defined as the absolute difference between the temporal distances of each country in a pair to Germany. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Adjusted p-values for the regressions in Table 1

	Dependent variable: Absolute difference in...			
	Risk taking	Prosociality	Neg. reciprocity	Patience
Column in Table 1	(3)	(5)	(7)	(9)
Unadjusted p-value	0.036	0.001	0.073	0.075
Adjusted p-value	0.058	0.013	0.060	0.060

E Robustness Check for Within-Country Analysis

Table 12: Preferences and temporal distance within country including observations with few respondents

	Dependent variable: Absolute difference in...									
	All preferences		Risk taking		Prosociality		Neg. reciprocity		Patience	
	Raw	Residual	Raw	Residual	Raw	Residual	Raw	Residual	Raw	Residual
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Temporal distance	0.097***	0.11***	0.051	0.067*	0.030	0.046**	0.049*	0.057**	0.089***	0.086***
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Country of residence FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country of birth FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6232	6232	6213	5995	6175	5958	6108	5890	6165	5947
R^2	0.32	0.33	0.25	0.26	0.27	0.28	0.26	0.25	0.20	0.22

Notes. OLS estimates, twoway-clustered standard errors (clustered at both countries of origin) in parentheses. The sample includes all population pairs, i.e., also those that consist of only one respondent each. The unit of observation is a population pair, which is defined as two groups who currently reside in the same country, but were born in different countries. The absolute difference in residual preferences is computed after individual-level preferences are residualized from age, age squared, gender, log household income p/c, educational attainment fixed effects, and marital status fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F Evidence Across Linguistic Groups

F.1 Overview of Interview Languages by Country

Afghanistan : Dari, Pashto
Algeria : Arabic
Argentina : Spanish
Australia : English
Austria : German
Bangladesh : Bengali
Bolivia : Spanish
Brazil : Portuguese
Cambodia : Khmer
Cameroon : English, French, Fulfulde
Canada : English, French
Chile : Spanish
China : Chinese
Colombia : Spanish
Costa Rica : Spanish
Croatia : Croatian
Czech Republic : Czech
Egypt : Arabic
Estonia : Estonian, Russian
Finland : Finnish
France : French
Georgia : Georgian, Russian
Germany : German
Ghana : Dagbani, English, Ewe, Twi
Greece : Greek
Guatemala : Spanish
Haiti : Creole
Hungary : Hungarian
India : Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Odia, Punjabi, Tamil, Telugu
Indonesia : Bahasa Indonesia
Iran : Farsi
Iraq : Arabic, Kurdish
Israel : Arabic, Hebrew

Italy : Italian
Japan : Japanese
Jordan : Arabic
Kazakhstan : Kazakh, Russian
Kenya : English, Swahili
Lithuania : Lithuanian
Malawi : Chichewa, English, Tumbuka
Mexico : Spanish
Moldova : Romanian, Russian
Morocco : Berber, French, Moroccan Arabic
Netherlands : Dutch
Nicaragua : Spanish
Nigeria : English, Hausa, Igbo, Pidgin English, Yoruba
Pakistan : Urdu
Peru : Spanish
Philippines : Bicol, Cebuano, Filipino, Hiligaynon, Iluko, Maguindanaon, Waray
Poland : Polish
Portugal : Portuguese
Romania : Romanian
Russia : Russian
Rwanda : English, French, Kinyarwanda
Saudi Arabia : Arabic
Serbia : Serbian
South Africa : Afrikaans, English, Sotho, Xhosa, Zulu
South Korea : Korean
Spain : Spanish
Sri Lanka : Sinhala, Tamil
Suriname : Dutch
Sweden : Swedish
Switzerland : French, German, Italian
Tanzania : English, Swahili
Thailand : Thai
Turkey : Turkish
Uganda : Ateso, English, Luganda, Runyankole
Ukraine : Russian, Ukrainian
United Arab Emirates : Arabic
United Kingdom : English

United States : English, Spanish
Venezuela : Spanish
Vietnam : Vietnamese
Zimbabwe : English, Ndebele, Shona

F.2 Residualized Preferences as Dependent Variables

Table 13: Residual preferences and linguistic distance across linguistic groups

	Dependent variable: Absolute difference in residualized...									
	All preferences		Risk taking		Prosociality		Neg. reciprocity		Patience	
Sample:	Full	Restricted	Full	Restricted	Full	Restricted	Full	Restricted	Full	Restricted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Linguistic distance	0.14***	0.14***	0.18***	0.18***	0.072*	0.075*	0.027	0.034	0.019	0.019
	(0.04)	(0.04)	(0.06)	(0.06)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)
Language FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3321	3081	3321	3081	3321	3081	3321	3081	3321	3081
R^2	0.56	0.54	0.62	0.62	0.47	0.45	0.51	0.50	0.40	0.40

Notes. OLS estimates, twoway-clustered standard errors (clustered at both languages in a pair) in parentheses. The unit of observation is a language pair. The dependent variables are absolute differences in language-level preferences, after individual-level preferences have been residualized from age, age squared, gender, log household income p/c, educational attainment fixed effects, and marital status fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: Average preferences and migratory distance to Ethiopia

	Dependent variable: Average ...							
	Risk taking		Prosociality		Patience		Neg. reciprocity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migratory Distance from Ethiopia	-0.016* (0.01)	0.010 (0.01)	0.0095 (0.01)	-0.0015 (0.02)	0.0019 (0.01)	0.027* (0.01)	-0.011 (0.01)	-0.012 (0.01)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	74	74	74	74	74	74	74	74
R^2	0.038	0.298	0.009	0.266	0.000	0.307	0.023	0.164

Notes. OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G Monotonic Selective Migration?

G.1 Baseline Analysis

In order to test whether monotonic selective migration causes the association between temporal distance and contemporary differences in preference levels, we relate the level of each preference in a population to the distance of that population to Ethiopia on the migratory path. Thus, we estimate the following equation:

$$\text{pref}_i = \alpha + \beta \times \text{migratory dist. from Ethiopia}_i + \gamma \times x_i + \epsilon_i$$

where pref_i is the average trait in country i, x_i is a vector of covariates, and ϵ_i a disturbance term. Note that this regression does *not* constitute a special case of the bilateral migratory distance regressions discussed in Section 4, because here the dependent variable is the *level* of a given preference, rather than the absolute difference to East Africa, i.e., Ethiopia. Thus, the regressions estimated above do not imply any prediction on the sign or significance of β.[20]

As Table 14 shows, contemporary preference levels are generally not associated with migratory distance to East Africa.

G.2 Dispersion of the Preference Pool

As mentioned in Section 7, the correlation between temporal distance and preference differences could be driven by a monotonic decrease of the *dispersion* of the preference pool along the migratory path, akin to a serial founder effect in population genetics.

[20]A special case of the general bilateral regression framework estimated in Section 4 would be

$$|\text{pref}_i - \text{pref}_{Ethiopia}| = \alpha + \beta \times \text{migratory dist. from Ethiopia}_i + \gamma \times |x_i - x_{Ethiopia}| + \epsilon_i$$

Since Ethiopia is not included in the Global Preferences Survey, we cannot estimate this equation.

Table 15: Preference dispersion and migratory distance from Ethiopia

	Dependent variable: Standard deviation in ...							
	Risk taking		Prosociality		Patience		Neg. reciprocity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migratory Distance from Ethiopia	-0.0051* (0.00)	0.0019 (0.00)	0.0042 (0.00)	0.0081 (0.01)	0.0078* (0.00)	0.0064 (0.01)	0.0033 (0.00)	0.0025 (0.01)
Risk taking		-0.0083 (0.04)						
Prosociality				-0.039 (0.04)				
Patience						0.32*** (0.04)		
Negative reciprocity								-0.19*** (0.04)
Continent FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	74	74	74	74	74	74	74	74
R^2	0.045	0.100	0.030	0.194	0.034	0.644	0.014	0.338

Notes. OLS estimates, robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

If the dispersion of the preference pool decreased monotonically along the migratory path, differences in preferences between later founder populations would mechanically be smaller than those between earlier founder population because the latter had larger variation in preferences to begin with.

To test whether the dispersion of the preference pool in a population is associated with its location on the migratory path, we relate the standard deviation of a population's preference trait to migratory distance from Ethiopia. We therefore estimate the following equation:

$$\text{sd\_pref}_i = \alpha + \beta \times \text{migratory dist. from Ethiopia}_i + \gamma \times x_i + \epsilon_i$$

where sd_pref_i is the standard deviation of the trait in country i, x_i is a vector of covariates, and ϵ_i a disturbance term. As Table 15 illustrates, the standard deviation in the different preference traits are generally unrelated to a population's migratory distance from Ethiopia.

H Definitions and Data Sources of Main Variables

H.1 Explanatory Variables

Fst genetic distance. Genetic distance between contemporary populations, taken from Spolaore and Wacziarg (2009) and Spolaore and Wacziarg (2017), respectively.

Linguistic distance (tree). Weighted linguistic distance between contemporary populations. Derived from the Ethnologue project data, taking into account all languages which are spoken by at least 5% of the population in a given country.

Linguistic distance (ASJP). Weighted linguistic distance between contemporary populations. Taken from `http://asjp.clld.org/`.

H.2 Covariates

Proportion female. Computed from the sociodemographic background data in the GPS.

Religious fractionalization. Index due to Alesina et al. (2003) capturing the probability that two randomly selected individuals from the same country will be from different religious / linguistic groups.

Percentage of European descent. Constructed from the "World Migration Matrix" of Putterman and Weil (2010).

Contemporary national GDP per capita. Average annual GDP per capita over the period 2001 – 2010, in 2005US$. Source: World Bank Development Indicators.

Democracy index. Index that quanties the extent of institutionalized democracy, as reported in the Polity IV dataset. Average from 2001 to 2010.

Colonial relationship dummies. Taken from the CEPII Geodist database at `http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=6`.

Geodesic distance, contiguity, longitude, latitude, area Taken from CEPII GeoDist database. The longitudinal distance between two countries is computed as

Longitudinal distance $= min\{|longitude_i - longitude_j|, 360 - |longitude_i| - |longitude_j|\}$

Suitability for agriculture. Index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH, taken from Michalopoulosa (2012).

Mean and standard deviation of elevation. Mean elevation in km above sea, taken from Ashraf and Galor (2013). Data originally based on geospatial elevation data reported by the G-ECON project (Nordhaus, 2006).

Precipitation. Average monthly precipitation of a country in mm per month, 1961-1990, taken from Ashraf and Galor (2013). Data originally based on geospatial average monthly precipitation data for this period reported by the G-ECON project (Nordhaus, 2006).

Temperature. Average monthly temperature of a country in degree Celsius, 1961-1990, taken from Ashraf and Galor (2013). Data originally based on geospatial average monthly temperature data for this period reported by the G-ECON project (Nordhaus, 2006).