<div align="center">

Online Appendix to Culture, Ethnicity and Diversity

by Klaus Desmet, Ignacio Ortuño-Ortín and Romain Wacziarg

April 2017

Appendix A - Theoretical Appendix

</div>

A.1. Proofs in the General Case

This Appendix generalizes the theoretical discussion of the indices of diversity presented in the main paper, to the case of q questions rather than one question. It provides proofs in this general case.

It is useful to start with some notation. A country is composed of n individuals characterized by the ethnic group to which they belong and by their cultural values or preferences. There are S ethnic groups, indexed by $s = 1, ..., S$. The share of each ethnic group in the population is w^s. Cultural values (or traits) are the answers to the q questions, each indexed by $i = 1, ..., q$. Each question i has $r(i)$ different possible answers, indexed by j. Focusing on a given country, w_{ij} is the share of the population that gives answer j to question i. Finally, w_{ij}^s is the share of individuals from ethnic group s that gives answer j to question i.

The type of an individual, k, is given by his ethnic group s and his answers to the q questions. We define a vector τ_k of dimension $1+q$ where the first component is a number from $\{1, 2, ..., S\}$ and denotes his ethnic group, and the remaining q components represent the answers to each of the q questions. For example, if there are two ethnic groups, $S = 2$, and three questions, $q = 3$, and each question has two answers, $r(i) = 2$, the vector $\tau_1 = \{1, 2, 1, 1\}$ characterizes the type of an individual (i.e. type 1) who belongs to the first ethnic group and who gives answers $2, 1, 1$ to the first, second and third question, respectively. Since we have a finite number of individuals, n, as well as a finite number of questions and answers, the total number of different types of individuals is finite. We denote by K the number of different types and by n_k the number of individuals of type k, so $\sum_{k=1}^{K} n_k = n$. The population share of individuals of type k is denoted by $f_k = n_k/n$, where of course $\sum_{k=1}^{K} f_k = 1$. We denote by $\xi(k, i)$ the answer given by an agent of type k to question i, and as $s(k)$ the ethnic group of a type k agent: $\tau_k = (s(k), \xi(k, 1), \xi(k, 2), ..., \xi(k, q))$.

A.1.1. - Proof that $v = CF$.

We start with the definition of the level of antagonism of an individual of type k when only cultural differences matter and we give the same weight to all the q questions:

$$v_k = 1 - \frac{1}{q} \sum_{i=1}^{q} w_{i, \xi(k,i)} \tag{1}$$

Summing over types and taking the population-weighted average:

$$v = \sum_{k=1}^{K} \left(1 - \frac{1}{q} \sum_{i=1}^{q} w_{i, \xi(k,i)} \right) \frac{n_k}{n}. \tag{2}$$

<div align="center">1</div>

Notice that (2) can be written as:

$$v = \sum_{k=1}^{K} \left(1 - \frac{1}{q} \sum_{i=1}^{q} w_{i,\xi(k,i)} \right) f_k = 1 - \frac{1}{q} \sum_{i=1}^{q} \sum_{k=1}^{K} w_{i,\xi(k,i)} f_k.$$

Next, note that:

$$\sum_{k=1}^{K} w_{i,\xi(k,i)} f_k = \sum_{j=1}^{r(i)} \sum_{k:\xi(k,i)=j} w_{i,\xi(k,i)} f_k = \sum_{j=1}^{r(i)} w_{ij} \sum_{k:\xi(k,i)=j} f_k = \sum_{j=1}^{r(i)} w_{ij}^2.$$

Thus, v becomes the following easy-to-calculate index of cultural fractionalization (CF):

$$CF = \frac{1}{q} \sum_{i=1}^{q} \left(1 - \sum_{j=1}^{r(i)} w_{ij}^2 \right). \tag{3}$$

A.1.2. Proof that $v = ELF$.

We start with the definition of the level of antagonism of an individual of type k when only ethnic differences matter:

$$v_k = 1 - w^{s(k)}. \tag{4}$$

Social antagonism, v, is the average of this probability over all individuals:

$$v = \sum_{k=1}^{K} (1 - w^{s(k)}) f_k. \tag{5}$$

It is easy to see that this is just the traditional index of ethnolinguistic fractionalization. Indeed, we have:

$$v = \sum_{k=1}^{K} (1 - w^{s(k)}) f_k = 1 - \sum_{s=1}^{S} \sum_{k:s(k)=s} w^{s(k)} f_k = 1 - \sum_{s=1}^{S} w^s \sum_{k:s(k)=s} f_k = 1 - \sum_{s=1}^{S} (w^s)^2. \tag{6}$$

Hence, v in this case becomes the common ELF index of ethnic fractionalization:

$$ELF = 1 - \sum_{s=1}^{S} (w^s)^2. \tag{7}$$

A.1.3. Proof that $v = \chi^2$.

Individual antagonism for question i and an agent of type k is given by

$$v_{ik} = \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}}$$

Averaging v_{ik} over all the q questions, giving the same weight to all of them, individual antagonism is:

$$v_k = \frac{1}{q} \sum_{i=1}^{q} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} \tag{8}$$

2

We add up the individual levels of antagonism across k, weighing by the population shares of each type k, to obtain social antagonism:

$$v = \frac{\sum_{k=1}^{K} v_k n_k}{n} = \frac{\sum_{k=1}^{K} \frac{1}{q} \sum_{i=1}^{q} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} n_k}{n} \tag{9}$$

Then:

$$v = \sum_{k=1}^{K} \frac{1}{q} \sum_{i=1}^{q} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} f_k = \frac{1}{q} \sum_{i=1}^{q} \sum_{k=1}^{K} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} f_k$$

Recall that the χ^2 index is given by:

$$\chi^2 = \frac{1}{q} \sum_{i=1}^{q} \sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w^s (w_{ij} - w_{ij}^s)^2}{w_{ij}}$$

Thus, it is enough to show that:

$$\sum_{k=1}^{K} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} f_k = \sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w^s (w_{ij} - w_{ij}^s)^2}{w_{ij}}, \quad \text{for all } i = 1, 2, ..., q \tag{10}$$

For each question $i = 1, 2, ..., q$ the right-hand term in (10) can be written as:

$$\sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w^s (w_{ij} - w_{ij}^s)^2}{w_{i,j}} = \sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w^s (w_{ij} - w_{ij}^s)(w_{ij} - w_{ij}^s)}{w_{ij}} \tag{11}$$

$$= \sum_{s=1}^{S} \sum_{j=1}^{r(i)} w^s (w_{ij} - w_{ij}^s) - \sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w^s w_{ij}^s (w_{ij} - w_{ij}^s)}{w_{ij}} \tag{12}$$

$$= \sum_{s=1}^{S} \sum_{j=1}^{r(i)} w^s w_{ij} - \sum_{s=1}^{S} \sum_{j=1}^{r(i)} w^s w_{ij}^s - \sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w^s w_{ij}^s (w_{ij} - w_{ij}^s)}{w_{ij}} \tag{13}$$

$$= -\sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w^s w_{ij}^s (w_{ij} - w_{ij}^s)}{w_{ij}}$$

and the left-hand term in (10) can be written as:

$$\sum_{k=1}^{K} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} f_k = \sum_{s=1}^{S} \sum_{k:s(k)=s} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} f_k \tag{14}$$

$$= \sum_{s=1}^{S} \sum_{j=1}^{r(i)} \sum_{k:s(k)=s, \xi(k,i)=j} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} f_k$$

$$= \sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w_{ij}^s - w_{ij}}{w_{ij}} \sum_{k:s(k)=s, \xi(k,i)=j} f_k$$

Notice that $\sum_{k:s(k)=s, \xi(k,i)=j} f_k$ is the population share of individuals who belong to ethnic group s and give answer j, thus:

$$\sum_{k:s(k)=s, \xi(k,i)=j} f_k = w^s w_{ij}^s \tag{15}$$

3

and from (14) and (15) we have:

$$\sum_{k=1}^{K} \frac{w_{i,\xi(k,i)}^{s(k)} - w_{i,\xi(k,i)}}{w_{i,\xi(k,i)}} f_k = \sum_{s=1}^{S} \sum_{j=1}^{r(i)} (\frac{w_{ij}^s - w_{ij}}{w_{ij}}) w^s w_{ij}^s = - \sum_{s=1}^{S} \sum_{j=1}^{r(i)} \frac{w^s w_{ij}^s (w_{ij} - w_{ij}^s)}{w_{ij}}$$

Thus, equality (10) holds for all questions i.

A.2. Drawbacks of FST and Uses of FST in the Past Literature

Drawbacks of F_{ST}. While the F_{ST} index is very commonly used in population genetics, it does have some drawbacks, as explained for example in Jost (2008), Meirmans and Hedrick (2011) and Jakobsson et al. (2013). The most important drawback, outlined in Jost (2008), relates to the properties of the decomposition of within and between fractionalization in a context where these measures are bounded above by 1. To illustrate this potential problem, let us denote, for a given question, between-group fractionalization by D. Such between-group fractionalization is defined by subtracting within-group fractionalization CF^W from total fractionalization CF, i.e., $D \equiv CF - CF^W$. Thus, this approach relies upon the additive decomposition of total fractionalization, but CF^W and D are not independent because we always have that $D + CF^W \leq 1$.[1] This constraint implies that D declines with within-group fractionalization CF^W regardless of the degree of cultural differentiation of ethnic groups.

A numerical example is useful to illustrate this drawback. Suppose that there is just one question and two ethnic groups of the same size. The question has four possible answers, a, b, c and d. The vector of answers for individuals from the first ethnic group is $\{0.1, 0.9, 0, 0\}$, i.e., 10% of them answer a, and 90% answer b. For the second ethnic group the vector of answers is $\{0.9, 0.1, 0, 0\}$. It is easy to check that in this society, $F_{ST} = 0.64$. Suppose a second society where those two vectors of answers are $\{0.5, 0.5, 0, 0\}$ and $\{0, 0, 0.5, 0.5\}$. It is clear that in this society culture and ethnicity overlap more strongly than in the first society. However, in this case we have $F_{ST} = 0.33$. The reason is that the second society displays a much higher degree of within-group heterogeneity than the first (a high CF^W), which drives down F_{ST} in spite of the higher degree of overlap between culture and ethnicity.

Our first overlap measure, χ^2, is not subject to this drawback, but empirically it does not matter which index we use: while the χ^2 index comes out directly from our model of antagonism, empirically χ^2 and F_{ST} are almost perfectly correlated (in our sample the correlation is 98%).

Uses of F_{ST} in the past literature. A few papers have used F_{ST} to measure between-group cultural heterogeneity, but their goals and methods are very different from ours. Bell et al. (2009) study inter-group competition and analyze whether there is more scope for selection based on cultural traits rather than on genetic traits. They use the WVS to compute a cultural F_{ST} measure between 150 pairs of neighboring countries. They show that this measure is an order of magnitude larger than an analogous measure of F_{ST} based on genetic data, suggesting a greater scope for cultural rather than genetic selection. In contrast to our approach, they measure cultural heterogeneity between countries rather than between groups within

---

[1] Jost (2008, pp 4018) provides a complete explanation of this constraint and its implication: "Additive partitioning of heterozygosity does not produce pure within-and between-subpopulation components; it is an incomplete partitioning". In our case cultural fractionalization is a parallel concept to heterozygosity in population genetics.

countries. In another paper, Ross et al. (2013) compute a measure of cultural F_{ST} based on between-group variation in folktales across different European ethnic groups. Again they are interested in comparing patterns of cultural F_{ST} to those of genetic F_{ST}, finding some similarities between the two. In contrast to our work, neither of these papers is interested in using cultural F_{ST} to measure the degree of overlap between ethnicity and culture, or in understanding how cultural F_{ST} relates to overall cultural heterogeneity and overall ethnolinguistic diversity. Instead, their focus is on the importance of cultural F_{ST}, relative to genetic F_{ST}. Finally, Spolaore and Wacziarg (2009, 2016) use a genetic F_{ST} as a measure of intergenerational divergence in a wide range of human traits transmitted culturally or biologically, in order to estimate the effects of barriers between populations on political and economic outcomes. In contrast to our approach, their F_{ST} is based on neutral genes, not cultural attitudes, and it measures distance between countries rather than between ethnic groups within countries.

A.3. Relationship Between χ^2, F_{ST} and Conventional Measures of Cross-Cuttingness.

Our proposed indices of overlap capture how much someone's ethnolinguistic identity reveals about his culture or preferences. High values imply that ethnolinguistic and cultural cleavages are reinforcing, whereas low values imply that they are cross-cutting. Our indices are thus related to an existing literature in political science concerned with the measurement of cross-cutting cleavages, starting with Rae and Taylor (1970, chapter 4). In what follows we start by discussing the Rae and Taylor measure of cross-cuttingness, and we then discuss the similarities and differences with our indices of overlap.

Consider two cleavages. In our terminology the first cleavage could refer to ethnicity and the second to culture (defined on a single dimension for now - say on a generic question i from the WVS). Assume there are S ethnic groups and $r(i)$ cultural groups. Fractionalization on cleavage 1 is ELF and fractionalization on cleavage 2 is CF_i, as previously defined. In Rae and Taylor's definition, if all those from a given ethnic group are also in a given cultural group, cleavages are perfectly reinforcing. They define cross-cutting XC as the "proportion of all pairs of individuals whose two members are in the same group of one cleavage but in a different group of the other cleavage" (p. 92), and show that, for a large enough population, we can write:

$$XC_i = \sum_{j=1}^{r(i)} w_{ij}^2 + \sum_{s=1}^{S} (w^s)^2 - 2\sum_{j=1}^{r(i)}\sum_{s=1}^{S} \left(w^s w_{ij}^s\right)^2 = 2F_i^c - CF_i - ELF, \tag{16}$$

where

$$F_i^c = 1 - \sum_{j=1}^{r(i)}\sum_{s=1}^{S} \left(w^s w_{ij}^s\right)^2. \tag{17}$$

Heuristically, F_i^c is fractionalization computed over all groups defined by both ethnicity and culture (so, for example, if $S = 3$ and $r(i) = 4$, there are 12 distinct groups defined by heterogeneity in both ethnicity and the answer to question i): F_i^c measures the probability that two randomly chosen individuals answer question i in a different way *or* belong to a different ethnic group. It is a measure of fractionalization where belonging to a different ethnic group or to a different cultural group defines different groups of individuals symmetrically, with the same weight.[2] With the definition above, it can be easily seen that XC_i is indeed

---

[2] In contrast ELF gives no weight to cultural heterogeneity while CF_i gives no weight to ethnic heterogeneity.

the probability that two randomly chosen individuals in the population belong to the same group on one cleavage but to a different group on the other cleavage. XC_i can be averaged over questions i to obtain an overall index XC.

Intuitively, being a measure of cross-cutting cleavages, XC should be negatively correlated with the χ^2 index (as well as F_{ST}) which are measures of reinforcing cleavages. However, both types of indices are quite different conceptually. This was already noted by Rae and Taylor (1970) and further discussed in Selway (2010, 2011). To put the distinction in stark focus with a simple example, consider the following distribution of individuals over two ethnicities and three possible answers to a cultural question:

(Entries are # of people)	Answer 1	Answer 2	Answer 3
Ethnic group 1	2	2	2
Ethnic group 2	1	1	1

Here both χ^2 and F_{ST} are obviously zero: both groups have the same distribution of answers as the population overall, and the share of between-group variance in total variance is zero (i.e. there is no between-group variance). However, F_i^C is 0.815, ELF is 0.444 and CF_i is 0.667, so that XC_i equals 0.518. As expected, a low value of χ^2 (or F_{ST}) corresponds to a high value of XC. When we double the size of ethnic group 1 proportionally for each possible answer, the values of χ^2 and F_{ST} are unchanged, as expected: they continue to be zero, since doubling the size of ethnic group 1 does not affect how informative ethnicity is about culture. In contrast, XC_i increases from 0.518 to 0.560, because the probability of two individuals sharing one cleavage but not the other increases. As this example illustrates, XC is sensitive to changes in group sizes that are not associated with changes in the degree to which ethnicity is informative about a person's cultural attitudes, the concept we have sought to capture in this paper so far. In fact, the example shows that XC has different properties from χ^2, an index that is derived directly from a simple model of social antagonism.

In addition to empirically analyzing χ^2 and F_{ST}, we also calculated the XC index of Rae and Taylor (1970) using our data. The correlations of XC with χ^2 and F_{ST} were, respectively, -0.43 and -0.45. Moreover, when using XC in our conflict regressions, instead of χ^2 or F_{ST}, we found that XC has a negative effect on the probability of civil war, as expected, but was not robustly significant across specifications corresponding to those in Table 5. However, as already noted, XC has very different properties from our proposed indices and does not come out of our model of antagonism. Since we have only provided a motivation for the χ^2 index in our theory, in the paper we focus mainly on χ^2 and F_{ST} (the latter being perhaps easier to interpret and very highly correlated with χ^2).

A.4. Distance-based Indices and Polarization

This Appendix proposes a number of additional diversity measures to analyze the link between culture and ethnicity. First, it shows how to introduce distances to compute cultural fractionalization and overlap measures. Second, it introduces measures of polarization, instead of fractionalization.

A.4.1. Introducing Cultural Distances

Consider one question from the WVS. There are r possible answers, indexed by i, and there are S ethnic groups, indexed by s. Let w_i denote the share of individuals in the entire population that chooses answer i, let w_i^s denote the share of individuals of group s that chooses answer i, and let w^s denote the share of individuals in the population that belongs to ethnicity s. Suppose that we can define a distance between the r possible answers to the question. Denote by d_{ij} the distance between answer i and answer j.

Once we take into account distances d_{ij} between responses i and j, our CF index can be readily extended to take into account those cultural distances by using Greenberg's B index:

$$CF_D = \sum_{i=1}^{r}\sum_{j=1}^{r} w_i w_j d_{ij}.$$

This index measures the expected distance between the answers given by two randomly picked individuals.

With Greenberg's index of cultural fractionalization in hand, it is easy to incorporate distances into our F_{ST} index:

$$\Phi_{ST} = \frac{CF_D - CF_D^W}{CF_D},$$

where $CF_D^W = \sum_{s=1}^{S} w^s CF_D^s$ and $CF_D^s = \sum_{i=1}^{r}\sum_{j=1}^{r} w_i^s w_j^s d_{ij}$. In population genetics this is often referred to as the "index of genetic differentiation", whereas F_{ST} is referred to as the "fixation index".

A.4.2. Polarization

In this section we describe the polarization counterparts of our fractionalization indices. We once again differentiate between indices without distances and indices with distances.

A.4.2.1. Polarization Indices without Cultural Distances. We can use the definition of polarization from Reynal-Querol (2002) or from Duclos, Esteban and Ray (2004) to come up with indices of cultural polarization and ethnic polarization.

- Cultural Polarization Measure:

$$CP = \sum_{i=1}^{r} w_i^{1+\alpha}(1 - w_i)$$

  where we use $\alpha = 1$.

- Ethnic Polarization Measure:

$$ELP = \sum_{s=1}^{S} (w^s)^{1+\alpha}(1 - w^s)$$

  where we again use $\alpha = 1$.

- Overlap Polarization Measure: For a polarization version of the overlap between culture and ethnicity we also follow Duclos, Esteban and Ray (2004). They analyze a situation where identification has to do with ethnicity, whereas alienation has to do with income differences. The difference here is that we

7

have ethnicity and culture, rather than ethnicity and income. Consistent with their equation (17), we can define the overlap between culture and ethnicity as:

$$P^2 = \sum_{s=1}^{S}\sum_{k \neq s}^{S}\sum_{i=1}^{r}\sum_{j \neq i}^{r}(w^s w_i^s)^2(w^k w_j^k).$$

Using the identification-alienation framework, here identification depends on $w^s w_i^s$, i.e., on the number of people of your ethnic group who give the same answer as you do divided by the society's total population. However, antagonism depends on $w^k w_j^k$, the number of people in another ethnic group who give a different answer divided by the society's total population. There is thus no antagonism between members of the same ethnic group or between individuals of different ethnic groups who choose the same answer. This is a natural way to model the "overlap" between ethnicity and culture within the polarization framework.

A.4.2.2. Polarization Indices with Cultural Distances.
The index of cultural polarization with distances can be written as:

$$CP_D = \sum_{i=1}^{r}\sum_{j=1}^{r} w_i^{1+\alpha} w_j d_{ij}.$$

This is equivalent to the index in equation (14) of Duclos, Esteban and Ray (2004).

Likewise, the measure of the overlap between culture and ethnicity is a generalization of P^2 that incorporates distances d_{ij}:

$$P_D^2 = \sum_{s=1}^{S}\sum_{k=1}^{S}\sum_{i=1}^{r}\sum_{j=1}^{r}(w^s w_i^s)^2(w^k w_j^k)d_{ij}.$$

A.5. A Simple Model of Ethnic and Cultural Conflict

In the paper, we assume that conflict is a monotonic function of the total level of antagonism, but we do not explicitly model this link. The objective of this Appendix is to present a simple model that rationalizes the link between antagonism and conflict. The main idea is standard in the literature on conflict (see, for example, Esteban and Ray, 2012). Groups invest resources in conflict and the probability of victory of group i is given by its share in total resources. Our focus here is on the resources invested in fighting, and not *per se* on conflict onset, incidence and intensity.

Suppose that there are only two ethnic groups and two cultural values, i.e., one question with two possible answers. We assume that the population of each ethnic group is 1, so that total population is 2. The results do not depend on this assumption. As in the main text of the paper, superscripts denote groups and subscripts cultural positions (answers). Denote by w_j^i, $i = 1, 2$, $j = 1, 2$, the number of individuals in ethnic group i with cultural position j (i.e. answer j). We then have $w_1^1 + w_2^1 = 1$ and $w_1^2 + w_2^2 = 1$. We write $w_j = w_j^1 + w_j^2$, $j = 1, 2$. Thus w_j is the total number of individuals with cultural position j. Without loss of generality we assume that $w_1^1 \geq w_2^1$. We focus on the case where ethnicity and culture are not independent so that, without loss of generality, we also assume that $w_1^1 > w_1^2$.

In the next three sections we will see that depending on the relevant type of cleavage, total resources invested in conflict are proportional to the corresponding index (ELF, CF, or χ^2). This result mirrors our finding that antagonism depends on these indices under similar assumptions about the cleavages that create antagonism.

A.5.1. The Overlap Channel

In this case we assume that conflict is always between ethnic groups (although the incentives of individuals would depend on their cultural positions). Since there is no income in our model civil conflict is always about ethnicity and culture differences. We assume that all the free rider problems within each group are solved. Moreover, there exists a "representative" member of the ethnic group who decides how much the group invests in conflict. This representative agent tries to maximize the average utility of the members of the group.

If group i wins, the "cultural policy" implemented will be the cultural position of the winning ethnic group. For example, if group 2 wins, the "effective" share of individuals with cultural position 1 in the whole society becomes w_1^2 and the share with cultural position 2 becomes w_2^2. An individual's payoff after the conflict is the effective share of people with his position after the conflict relative to the share with his position before the conflict. Hence, if group 2 wins, an individual of group 1 with position 1 has payoff $\frac{w_1^2}{w_1}$; and if group 1 wins, his payoff is $\frac{w_1^1}{w_1}$. The antagonism experienced by an individual of group 1 with position 1 is then defined as the difference between the payoff when his own group wins and the actual payoff. Therefore, if group 2 wins, an individual of group 1 with position experiences antagonism $\frac{w_1^1-w_1^2}{w_1}$, whereas if group 1 wins, he experiences antagonism $\frac{w_1^1-w_1^1}{w_1} = 0$.

If group 1 spends per capita resources r^1 and group 2 spends per capita resources r^2, the probability that group 1 wins is:

$$p = \frac{w^1 r^1}{w^1 r^1 + w^2 r^2} \tag{18}$$

We write the total resources spent by the two groups as $R^1 = w^1 r^1$ and $R^2 = w^2 r^2$. We identify the intensity of the conflict, R, with the total resources spent on the conflict $R = R^1 + R^2$.

The expected cost of conflict for an ethnic group is the expected average animosity plus the resources invested in conflict. Thus, if group 1 spends r^1, its expected cost is:

$$EC^1 = w_1^1 \left(p \times 0 + (1-p)\frac{w_1^1 - w_1^2}{w_1} \right) + w_2^1 \left(p \times 0 + (1-p)\frac{w_2^1 - w_2^2}{w_2} \right) + r^1 \tag{19}$$

In the same way, the cost for group 2 is

$$EC^2 = w_1^2(p\frac{w_1^2 - w_1^1}{w_1} + (1-p) \times 0) + w_2^2(p\frac{w_2^2 - w_2^1}{w_2} + (1-p) \times 0) + r^2 \tag{20}$$

Note that the "cultural minority" in, say, ethnic group 1 prefers the victory of ethnic group 2 because $w_2^1 - w_2^2 < 0$. They, however, cannot collaborate with the other ethnic group. That is, groups go to war about culture, but war has to be waged along ethnic lines. This illustrates how culture and ethnicity overlap in this type of conflict. Indeed, if members of the cultural minority of one ethnic group could collaborate with like-minded people of the other ethnic group, the conflict would be just cultural – this possibility is analyzed below.

Our equilibrium concept is Nash. Each group chooses r^i that minimizes the expected cost, taking the other group's choice as given. The following shows that the intensity of conflict is proportional to the χ^2 index as defined in the main text of the paper.

Lemma 1 *The intensity of conflict is proportional to* χ^2. *Namely:*

$$R = \frac{\chi^2}{2}$$

Proof. We can write the expected cost function (19) as

$$(1-p)A + r^1 \tag{21}$$

where

$$A = \frac{w_1^1}{w_1}(w_1^1 - w_1^2) + \frac{w_2^1}{w_2}(w_2^1 - w_2^2). \tag{22}$$

Minimizing (21) is equivalent to minimizing $pA + r^1$, where p is given by (18). The first order condition is

$$\frac{w^1 R - R^1 w^1}{(R)^2} A = 1$$

or

$$\frac{w^1 R^2}{(R)^2} A = 1. \tag{23}$$

In the same way, for group 2, (20) can be written as

$$pB + r^2,$$

where

$$B = \frac{w_1^2}{w_1}(w_1^2 - w_1^1) + \frac{w_2^2}{w_2}(w_2^2 - w_2^1) \tag{24}$$

and the corresponding first order condition is

$$\frac{w^2 R^1}{(R)^2} B = 1. \tag{25}$$

It's not difficult to show that in our case

$$\chi^2 = \frac{(w_2^1 w_1^2 - w_1^1 w_2^2)^2}{(w_1^1 + w_2^1)(w_1^1 + w_1^2)(w_2^1 + w_2^2)(w_1^2 + w_2^2)} = \frac{(w_2^1 w_1^2 - w_1^1 w_2^2)^2}{(w_1^1 + w_1^2)(w_2^1 + w_2^2)}$$

We can manipulate (22) to obtain

$$A = \frac{(w_2^1 w_1^2 - w_1^1 w_2^2)^2}{(w_1^1 + w_2^1)^2} \frac{(w_2^1 w_1^2 - w_1^1 w_2^2)^2}{(w_1^1 + w_1^2)(w_2^1 + w_2^2)(w_1^2 + w_2^2)} = \chi^2$$

In the same way we can show that

$$B = \chi^2$$

From (23) and (25) we have

$$\frac{R^2 \chi^2}{(R)^2} = 1 \tag{26}$$

10

and

$$\frac{R^1\chi^2}{(R)^2} = 1 \tag{27}$$

and these two equations together imply

$$R^1 = R^2$$

and using (26) again

$$\frac{R^1\chi^2}{(2R^1)^2} = 1$$

so that

$$R^1 = \frac{\chi^2}{4}$$

and

$$R = R^1 + R^2 = \frac{\chi^2}{2}$$

∎

From this we can conclude that if war is about culture but is waged along ethnic lines, the intensity of conflict will be proportional to χ^2. Note that when there is no overlap between culture and ethnicity, $w_1^1 = w_1^2$ and $w_2^1 = w_2^2$. In that case $\chi^2 = 0$ and there is no conflict.

A.5.2. The Ethnic Heterogeneity Channel

In this case culture plays no role; the only thing that matters is ethnicity. If ethnic group i loses, the antagonism it experiences is 1, and if it wins, its antagonism is zero. Thus, the expected per capita cost for group 1 is

$$EC^1 = p \times 0 + (1-p)1 + r^1 \tag{28}$$

and the expected per capita cost for group 2 is

$$EC^2 = p \times 1 + (1-p)0 + r^2 \tag{29}$$

where

$$p = \frac{w^1 r^1}{w^1 r^1 + w^2 r^2}.$$

The first order condition for group 1 is

$$\frac{w^1(w^1 r^1 + w^2 r^2) - r^1 w^1 w^1}{(w^1 r^1 + w^2 r^2)^2} = \frac{w^1 w^2 r^2}{(w^1 r^1 + w^2 r^2)^2} = 1 \tag{30}$$

and the first order condition for group 2 is

$$\frac{w^1 w^2 r^1}{(w^1 r^1 + w^2 r^2)^2} = 1. \tag{31}$$

From (30) and (31) we have

$$r^1 = r^2 = r.$$

Hence, from (31) we obtain that

$$r = \frac{w^1 w^2}{2}.$$

11

Total resources spent in the conflict are then

$$R = w^1 r + w^2 r = w^1 w^2 = \frac{ELF}{2}.$$

(32)

From this we can conclude that if conflict has to do with ethnicities *per se* and not with culture, then the intensity of conflict will be proportional to the ethnic ELF measure.

A.5.3. The Cultural Heterogeneity Channel

Here ethnicity plays no role in conflict; the only thing that matters are cultural differences. By analogy with the previous argument, conflict intensity will be proportional to cultural fractionalization. The only difference in the proof is that now we have two cultural groups of sizes w_1 and w_2, instead of two ethnic groups of sizes w^1 and w^2.

Appendix B - Empirical Appendix[3]

B.1. Ethnicity and Culture in Other Surveys

B.1.1. Afrobarometer.

We conduct an analysis similar to that in Section 2 using the 5th wave of Afrobarometer (2008). The wave covers 34 Sub-Saharan African countries, of which 28 have ethnicity data and are thus retained. There are 192 questions on values, norms and preferences in this survey, 18 admitting a binary response, 159 where the response is on an ordered scale and 15 which admit more than two unordered responses. These 15 questions were converted into a set of 97 binary questions in the same way as was done for the WVS/EVS, resulting in a total of 274 available questions for us to analyze. A specificity of Afrobarometer is that most questions were asked in every country, so there is little heterogeneity in the set of questions used (the mean number of questions asked in each of the 28 surveyed countries is 272, with a standard deviation of 1.6).

For ethnic identity, we relied entirely on the classification provided by Afrobarometer (variable Q84: "What is your tribe or ethnic group?"). The number of ethnic groups varies by country, ranging from 3 to 61. Finally, the regression specification includes as controls the respondent's age (Q1), gender (Q101), education of respondent (Q97), urban/rural status of respondent (URBRUR) and present living conditions (Q3B) as a proxy for income.

Results obtained using the Afrobarometer survey are presented in Appendix Table B2. Confirming results for Sub-Saharan Africa from the WVS/EVS, the share of questions for which ethnicity significantly predicts responses is high, on average 57% (versus 62% in the WVS/EVS). This average masks interesting variation across countries. Some countries like South Africa and Nigeria display very high shares (86% and 85%, respectively), while some small countries such as Lesotho and Cape Verde display much smaller shares

---

[3]All the data and programs used to compute the results contained in this Empirical Appendix are available from the authors upon request.

(respectively 8% and 12%).[4] Interestingly, the results for South Africa and Nigeria mirror those from the WVS/EVS, where the corresponding shares were 88% and 84%, despite different survey questions, ethnic nomenclature, methods and samples.

B.1.2. Latinobarómetro.

We also use the Latinobarómetro, which covers Latin American countries. The survey asks a question on ethnic identity only since 2007, so we are constrained to using waves for 2007, 2008, 2009 and 2010, covering 18 countries. The survey classifies questions into categories, and we focus on the categories that refer to values, opinions, and preferences. Among those, however, a few questions about the factual situation of the respondent crept in, and we removed them from consideration. We ended up with 231 binary response questions, 358 questions for which the response is on an ordered scale, and 19 questions with more than two unordered response categories. The latter were transformed into a series of 96 binary response questions, as before. We ended up with a total of 685 usable questions. As was the case for the Afrobarometer, the availability of questions across countries did not vary much. The average country had 678 questions (with a standard deviation of less than 2).

Ethnic identity is as defined by Latinobarómetro, and represents a type of classification very similar to the one we used for Latin American countries in the WVS/EVS. However, the group shares are very different, and appear non-represeantative of the actual group shares of various ethnic groups in Latin American countries. There are seven ethnic categories, corresponding to the respondent's race (the variable coding ethnicity is named A505206). The categories are Asian, Black, Indigenous, Mestizo, Mulatto, White and other race. The regression specification includes ethnic dummies, sex (S01), age (S02), education (S51) and socioeconomic level (S62) as a proxy for income.

Results obtained using the Latinobarómetro survey are presented in Appendix Table B3. The share of questions for which ethnicity dummies are jointly significant correlates of answers is 32.5%. As in the WVS/EVS survey, this average does not seem to vary much across question categories or question type (although, as before, the share is slightly smaller for binary questions created from underlying unordered multiple response questions). There appears to be only limited variation across countries. Argentina displays the lowest share in this sample (17.8%) while the Dominican Republic displays the largest (60%), but most shares are comprised in a tight band between 20% and 40%. The relatively low share obtained with Latinobarómetro is in line with results obtained previously using the WVS/EVS.

B.2. Endogeneity of CF and χ^2 in the Civil Conflict Regressions.

Endogeneity Concerns. As in most of the literature on civil conflict, in the main paper we treated our heterogeneity measures as exogenous to conflict. As long as we limit attention to ethnolinguistic fractionaliza-

---

[4]For Lesotho, the ethnic nomenclature provided by Afrobarometer is at the level of clans. Lesotho is considered by ethnographers to be a very ethnically homogeneous country (99.7% of the population belongs to the Sotho ethnic group, according to the CIA World Factbook, 2009). This homogeneity is reflected in the fact that different clans do not seem to differ much in terms of values, norms and preferences. Despite the questionable ethnic classification proposed for Lesotho by the Afrobarometer, we chose to leave this country in our sample. Excluding Lesotho from the Afrobarometer sample raises the average share of regressions where ethnicity significantly predicts culture to 59%.

tion, and include a suitably expansive set of controls, this can be justified as ethnolinguistic fractionalization is very time-persistent and is unlikely to change very much as a result of conflict. The same cannot necessarily be said of cultural heterogeneity and the overlap of culture and ethnicity.[5] The experience of civil wars can lead people to change their values and preferences, and respond differently to questions from the World Values Survey. This in turn can lead measures of cultural diversity such as CF to change as a result of conflict (though it is not clear *a priori* in what direction, hence the sign of the endogeneity bias is not clear). Similarly, civil war, especially when there is an ethnic component, can change the salience of ethnic identity, leading ethnic groups to adopt values and attitudes that differ more than they did prior to the war, or on the contrary can lead the victor to impose their values and preferences on the vanquished, therefore affecting the degree of overlap between ethnicity and culture (again in an ambiguous direction). This could be a problem particularly in our setting as the questions from the WVS used to characterize preferences and values are from survey waves from 1981 to 2008, while our main civil war dataset (from Fearon and Laitin, 2003) covers 1945 to 1999.

Dynamic Specification. We already partly address this potential endogeneity problem in the regressions shown so far as we adopt a dynamic specification for the incidence of civil wars, i.e., we include a term for lagged civil war on the right-hand side of the specification, in keeping with the usual practice in the literature (see in particular Fearon and Laitin, 2003, p. 84 and Esteban, Mayoral and Ray, 2012, p. 1318). Since civil war incidence is highly autocorrelated, this purges CF and χ^2 of much of their variation attributable to past wars.

Persistence of the Indices. Before discussing our strategies to further address endogeneity, we examine the persistence of the measures of diversity across waves. We compute correlations of CF and χ^2 across the different waves of the WVS/EVS. Appendix Table B33 reports the results (Appendix Table B47, columns (4) and (5) do something similar for our public goods results). The correlations in CF range from 0.68 to 0.88 and for χ^2 from 0.71 to 0.97. We note that the time persistence of our main predictor of conflict, χ^2, is quite high, reaching 0.97 from wave 3 to wave 4 and 0.95 from wave 4 to wave 5. Although these correlations are based on few observations – not many countries have consistent ethnolinguistic categories across waves – the persistence of the indices gives some confidence that the endogeneity of χ^2 due to conflict is not an overwhelming concern.

Additional Strategies to Deal with Endogeneity. We consider three additional approaches. First, we focus on questions that display a high degree of persistence in cultural fractionalization across waves, and are thus less likely than other questions to respond endogenously to external events such as civil wars. For each question i, country c and WVS wave w, consider cultural fractionalization CF_{icw}. For each question i and each country c we compute the coefficient of variation of CF_{icw} across waves w, and average this

---

[5] A related concern is whether the WVS/EVS is less likely to ask about an individual's ethnicity in countries that have suffered violent conflict. There are only nine countries for which the WVS/EVS does not allow us to construct ethnolinguistic categories (Austria, Belgium, Colombia, Denmark, Greece, Croatia, Ireland, Portugal and El Salvador). To see whether this introduces any bias, we can compare the probability of war in any given year in the sample we are using (14%) to the probability of war in the limited sample of nine countries (17%). We therefore find little scope or evidence that the inability to infer a respondent's ethnicity in some countries introduces selection bias.

coefficient of variation across countries for each question i. This gives a measure of persistence for each question i. We then remove from consideration every question with a coefficient of variation in excess of 0.1, which leaves us with about 60% of the questions used previously – the ones with the most time persistent value of CF_{icw}. We reran our baseline specification (the one in column 4 of Table 5) with CF and χ^2 computed from this restricted set of questions. The results are presented in column (1) of Appendix Table B34. The signs of our main effects remain the same, namely CF affects conflict negatively (albeit the effect is no longer statistically significant at the 5% level) and χ^2 affects conflict incidence positively and remains significant at the 5% level. The magnitude of the effect falls slightly in standardized terms, with a one standard deviation increase in χ^2 associated with a 13% increase in the probability of conflict.

Our second approach is to limit our sample to the post-1970 period. The idea is that if endogeneity were a strong concern, we should find different results in this subsample compared to the full sample. The argument could take various forms. On the one hand, if we limit attention to wars that occur closer to the date when we observe values, there is perhaps greater potential for recent wars to affect values, and then in turn cultural diversity and χ^2. On the other hand, if the lag with which war may affect values is substantial, by focusing on a recent sample, values may not yet have had time to change, and therefore CF and χ^2 may not yet have changed in response to civil war. In either case, if reverse causality were a concern we would observe different estimates of the effect of CF and χ^2 in the post-1970 sample and in the whole sample. Column (3) of Appendix Table B34 presents the results, which are similar to the baseline regression of Column 4 of Table 5. Indeed, the standardized effect of χ^2 stands equal to the one previously calculated, at 17%. The standardized effect of CF on the probability of civil war also remains equal to -10%. These results have the added advantage of showing the stability and robustness of our estimated effects to the sample period under consideration.

Out third approach is a variation on the previous one. Here, we limit attention to respondents born before 1950 *and* to the post-1970 sample. The effects could once again go in a variety of directions, but the argument is again that the estimates would be *different* if endogeneity were a big concern. On the one hand, if respondents' values are formed in early adulthood and change little after that, since every respondent in the sample would be at least 20 years old in the event of a civil war, their cultural values may respond less than younger individuals to the event of a war. On the other hand, if one's view was that these individuals were the most likely to be affected by a civil war because they were most likely to be combatants or to be affected by the war in adult age, their values may be most likely to be affected by the war. Either way, war would affect values, and therefore potentially CF and χ^2 also (although, again, in unknown directions). Column (4) of Appendix Table B34 presents the results. The effects of CF and χ^2 on the probability of war have the same signs, although their magnitude is smaller than in the baseline.

Summary. While we do not want to place too much weight on any one of these tests in isolation, taken together they do suggest that our main results are quite stable when looking only at the post-1970 sample, when looking only at respondents born before 1950, and when including only questions for which question-by-question cultural diversity. CF_i is most stable across survey waves. Under reverse causality, if war had a strong causal effect on CF or χ^2, we would have expected estimates under each of these modifications to differ from the baseline. They do not differ greatly.

Appendix B3: Robustness Tables for Section 2 Results

Appendix Table B1 - Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, Country by Country

Country	# of Regressions / Questions	Share of Regressions with Jointly Significant Ethnic Dummies	ΔR²	# of Ethnic Groups
Albania	350	0.109	0.209	3
Algeria	287	0.296	0.724	4
Andorra	285	0.379	1.081	5
Armenia	274	0.255	0.368	4
Australia	405	0.437	0.828	7
Azerbaijan	272	0.665	1.748	10
Bangladesh	256	0.141	0.389	4
Belarus	274	0.336	0.581	5
Bosnia and Herzegovina	346	0.676	1.700	4
Brazil	282	0.106	0.438	5
Bulgaria	267	0.352	0.937	4
Burkina Faso	282	0.585	1.844	9
Canada	396	0.639	1.130	3
Chile	349	0.186	0.451	6
Cyprus	277	0.650	3.279	5
Czech Republic	271	0.059	0.304	4
Dominican Republic	271	0.089	1.961	6
Egypt	226	0.624	0.843	5
Estonia	274	0.701	3.148	3
Ethiopia	281	0.712	2.024	8
Finland	403	0.231	0.511	5
France	193	0.233	1.053	6
Georgia	398	0.445	0.819	12
Germany	287	0.206	0.405	5
Ghana	254	0.748	2.014	6
Great Britain	198	0.369	1.432	7
Guatemala	291	0.210	0.241	2
India	398	0.990	6.256	16
Indonesia	365	0.770	2.083	9
Iran	242	0.603	0.929	9
Iraq	160	0.825	3.499	5
Israel	81	0.642	2.431	2
Jordan	235	0.149	0.635	7
Kyrgyzstan	287	0.551	1.731	4
Latvia	273	0.407	2.145	11
Lithuania	273	0.282	0.483	2
Macedonia	345	0.739	3.777	6
Malaysia	274	0.704	3.387	15
Mali	281	0.306	1.381	9

Country	# of Regressions / Questions	Share of Regressions with Jointly Significant Ethnic Dummies	ΔR²	# of Ethnic Groups
Mexico	406	0.259	0.495	6
Moldova	401	0.546	1.414	6
Morocco	386	0.215	0.513	6
New Zealand	267	0.199	0.858	6
Nigeria	320	0.838	1.410	5
Norway	285	0.225	0.281	2
Pakistan	198	0.697	1.824	8
Peru	245	0.269	0.784	7
Philippines	290	0.617	3.420	20
Poland	277	0.061	0.211	3
Puerto Rico	271	0.089	0.572	6
Romania	266	0.147	0.543	5
Russian Federation	322	0.522	0.611	4
Saudi Arabia	208	0.418	1.594	11
Serbia	276	0.370	1.401	7
Singapore	217	0.705	2.836	6
Slovakia	271	0.421	1.120	5
Slovenia	279	0.090	0.382	4
South Africa	447	0.884	2.004	12
Spain	431	0.548	1.140	5
Sweden	287	0.220	0.988	7
Switzerland	240	0.575	1.743	5
Taiwan	371	0.337	0.723	5
Tanzania	290	0.190	0.392	3
Thailand	287	0.976	5.064	7
Trinidad and Tobago	278	0.237	0.852	6
Turkey	280	0.493	1.306	5
Uganda	289	0.346	2.647	9
Ukraine	395	0.565	0.874	3
United States	345	0.368	0.675	6
Uruguay	398	0.106	0.486	6
Venezuela	217	0.078	0.523	6
Viet Nam	284	0.342	0.226	2
Zambia	280	0.782	3.481	18

Note: Using an alternative, more restricted set of questions, we find these shares to be remarkable stable. The correlation between the two series is 98.84%, despite using only half the questions, and leaving out the transformed multinomial questions.

Note that some countries have very small numbers because of a lopsided distribution of respondents across ethnic groups. Examples include Czech Republic, Poland, Slovenia, where there are few groups and a very small number of respondents in some groups.

Appendix Table B2 - Afrobarometer: Overall Results and Breakdown by Country

	# of Regressions	Share of Jointly Significant Ethnic Dummies	# of Ethnic Groups
Overall Sample	**7,610**	**0.570**	**-**
Binary response questions	486	0.638	-
Binary from unordered multiple response questions	2,695	0.425	-
Scale response questions	4,429	0.651	-
Benin	271	0.672	11
Botswana	272	0.272	26
Burkina Faso	272	0.419	21
Burundi	272	0.143	3
Cameroon	273	0.564	46
Cape Verde	269	0.123	9
Cote d'Ivoire	273	0.542	5
Ghana	272	0.757	24
Guinea	272	0.688	10
Kenya	273	0.897	23
Lesotho	273	0.084	16
Liberia	271	0.498	17
Madagascar	266	0.827	22
Malawi	271	0.668	13
Mali	271	0.469	18
Mauritius	271	0.435	8
Mozambique	274	0.832	25
Namibia	273	0.608	15
Niger	271	0.443	8
Nigeria	274	0.854	43
Senegal	271	0.262	10
Sierra Leone	272	0.496	15
South Africa	272	0.860	14
Tanzania	273	0.861	61
Togo	272	0.574	24
Uganda	273	0.934	29
Zambia	272	0.533	31
Zimbabwe	271	0.646	24

Notes: 28 countries. Based on data from the 5th Afrobarometer wave only (2014). Results for the 4th wave (20 countries only, 2009) are very similar and available in the previous version of this paper.

A breakdown by question category is not available for this dataset.

Regression specification includes ethnic dummies, age (Q1), present living conditions (Q4B) as a proxy for income, and gender of respondent (Q101).

Appendix Table B3 - Latinobarómetro: Overall Results and Breakdown by Question Category, Question Type and Country

	# of Regressions	Share of Jointly Significant Ethnic Dummies
Across All Questions and Countries:	**12,210**	**0.325**
Breakdown by Question Category:		
A: Democracy, Participation, Social Values, Trust	6,546	0.327
B: Public Policies, Corruption, Labor	3,497	0.326
C: Economics, Development, Entrepreneurship	1,099	0.354
D: Means of Communication	852	0.285
E: Political Developments	216	0.292
Breakdown by Question Type:		
Binary	4,132	0.309
Scale	6,350	0.367
Binary from Unordered Multiple Response Questions	1,728	0.213
Breakdown by Country		
Argentina	679	0.178
Bolivia	682	0.453
Brazil	677	0.236
Colombia	680	0.228
Costa Rica	676	0.249
Chile	677	0.297
Ecuador	680	0.415
El Salvador	679	0.432
Guatemala	680	0.296
Honduras	678	0.355
Mexico	679	0.199
Nicaragua	679	0.247
Panama	678	0.289
Paraguay	676	0.408
Peru	680	0.334
Uruguay	675	0.215
Venezuela	677	0.427
Dominican Republic	678	0.600

Notes: 18 countries. This is based on 2007, 2008, 2009, 2010 waves of Latinobarómetro, the only available waves where the ethnicity question was asked.

Regression specification includes ethnic dummies, sex (S01), age (S02), respondent education (S51) and socioeconomic level (S62) as a proxy for income.

Appendix Table B4 – Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Region

	# of regres-sions	Share of regressions w/ jointly significant ethnic dummies	R^2 without ethnic dummies	R^2 with ethnic dummies	ΔR^2
Whole sample	**20,207**	**0.435**	**2.679**	**4.081**	**1.402**
Africa	3,333	0.580	2.492	4.194	1.702
Of which: Sub-Saharan Africa	2,434	0.667	2.391	4.476	2.085
Of which: North Africa	899	0.344	2.766	3.430	0.663
Europe	7,492	0.384	3.050	4.181	1.131
Of which: Western and Southern Europe	2,369	0.313	3.567	4.399	0.832
Of which: Eastern and Central Europe	5,123	0.417	2.811	4.081	1.270
Asia	5,239	0.561	2.244	4.380	2.136
Of which: East and Southeast Asia	1,871	0.617	2.050	4.437	2.387
Of which: South Asia	654	0.657	2.387	6.347	3.959
Of which: Southwestern and Central Asia	1,511	0.479	2.223	3.391	1.168
Of which: Middle East	1,203	0.525	2.494	4.464	1.971
America	3,471	0.235	2.525	3.222	0.697
Of which: North America	741	0.513	3.157	4.075	0.918
Of which: Latin America and Caribbean	2,730	0.159	2.354	2.990	0.636
Oceania	672	0.342	3.669	4.509	0.840

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms.

This table replicates the baseline results in Table 1, <u>without controlling for regional dummies</u>, using the sample of 68 countries for which regional dummies are available (lost countries compared to baseline: Pakistan, Poland, Singapore, Trinidad and Tobago, and Tanzania).

Appendix Table B5 - Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	R^2 without ethnic dummies	R^2 with ethnic dummies	ΔR^2
Breakdown by Question Category					
A: Perceptions of Life	4,090	0.431	3.219	4.578	1.358
B: Environment	930	0.431	2.212	3.670	1.458
C: Work	2,278	0.398	2.369	3.525	1.156
D: Family	1,243	0.447	3.240	4.602	1.362
E: Politics and Society	8,529	0.416	2.413	3.751	1.338
F: Religion and Morals	2,160	0.520	3.312	5.061	1.749
G: National Identity	977	0.510	1.805	3.744	1.939
Breakdown by Question Type					
Binary questions	4,278	0.435	2.841	4.252	1.412
Binary from unordered response questions	6,628	0.363	1.613	2.719	1.105
Scale questions	9,301	0.487	3.365	4.973	1.608

Note: This result does not change if you break it down by continent: there is little difference in shares of questions with significant ethnolinguistic dummies when the breakdown by category is done continent by continent. R^2 is expressed in % terms.

This table replicates the baseline results in Table 2, <u>without controlling for regional dummies</u>, using the sample of 68 countries for which regional dummies are available (lost countries compared to baseline: Pakistan, Poland, Singapore, Trinidad and Tobago, and Tanzania).

Appendix Table B6 – Joint Significance of Regional Dummies in Questions from the World Values / European Values Integrated Surveys, by Region

	# of regres-sions	Share of regressions w/ jointly significant region dummies	R^2 without region dummies	R^2 with region dummies	ΔR^2
Whole sample	20,238	0.746	2.633	6.231	3.597
Africa	3,341	0.833	2.512	6.458	3.946
Of which: Sub-Saharan Africa	2,435	0.842	2.405	6.344	3.939
Of which: North Africa	906	0.809	2.801	6.763	3.963
Europe	7,499	0.698	3.007	5.882	2.876
Of which: Western and Southern Europe	2,374	0.575	3.459	6.062	2.603
Of which: Eastern and Central Europe	5,125	0.754	2.797	5.799	3.002
Asia	5,244	0.860	2.203	7.629	5.426
Of which: East and Southeast Asia	1,871	0.785	2.050	5.815	3.765
Of which: South Asia	654	0.995	2.151	11.951	9.800
Of which: Southwestern and Central Asia	1,511	0.919	2.241	7.793	5.552
Of which: Middle East	1,208	0.831	2.421	7.894	5.472
America	3,482	0.677	2.424	4.978	2.554
Of which: North America	746	0.634	3.096	4.511	1.415
Of which: Latin America and Caribbean	2,736	0.689	2.241	5.105	2.864
Oceania	672	0.326	3.511	4.566	1.055

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms.

This table replicates the baseline results in Table 1 <u>doing the exercise for regional dummies instead of ethnic dummies</u>, using the sample of 68 countries for which both ethnic and regional dummies are available (lost countries compared to baseline: Pakistan, Poland, Singapore, Trinidad and Tobago, and Tanzania).

Appendix Table B7 - Joint Significance of Regional Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type

	# of regres-sions	Share of regressions with jointly significant region dummies	R^2 without region dummies	R^2 with region dummies	ΔR^2
Breakdown by Question Category					
A: Perceptions of Life	4,090	0.727	3.148	6.469	3.322
B: Environment	932	0.745	2.178	6.629	4.451
C: Work	2,280	0.729	2.324	5.413	3.088
D: Family	1,243	0.770	3.216	6.816	3.600
E: Politics and Society	8,554	0.742	2.370	5.915	3.546
F: Religion and Morals	2,162	0.795	3.289	7.389	4.100
G: National Identity	977	0.771	1.751	6.211	4.459
Breakdown by Question Type					
Binary questions	4,280	0.717	2.817	6.284	3.467
Binary from unordered response questions	6,657	0.689	1.577	4.456	2.879
Scale questions	9,301	0.801	3.305	7.476	4.171

Note: This result does not change if you break it down by continent: there is little difference in shares of questions with significant ethnolinguistic dummies when the breakdown by category is done continent by continent. R^2 is expressed in % terms.

This table replicates the baseline results in Table 2 doing the exercise for regional dummies instead of ethnic dummies, using the sample of 68 countries for which both ethnic and regional dummies are available (lost countries compared to baseline: Pakistan, Poland, Singapore, Trinidad and Tobago, and Tanzania).

Appendix Table B8 – Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Region, controlling for region dummies

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	R^2 without ethnic dummies	R^2 with ethnic dummies	ΔR^2
Whole sample	20,207	0.313	6.347	7.211	0.864
Africa	3,333	0.354	6.170	7.193	1.023
Of which: Sub-Saharan Africa	2,434	0.404	6.216	7.444	1.228
Of which: North Africa	899	0.219	6.045	6.512	0.467
Europe	7,492	0.312	5.783	6.617	0.834
Of which: Western and Southern Europe	2,369	0.255	6.339	7.031	0.692
Of which: Eastern and Central Europe	5,123	0.338	5.526	6.425	0.899
Asia	5,239	0.374	8.225	9.243	1.018
Of which: East and Southeast Asia	1,871	0.435	5.816	7.137	1.321
Of which: South Asia	654	0.433	16.632	17.520	0.887
Of which: Southwestern and Central Asia	1,511	0.317	7.617	8.289	0.673
Of which: Middle East	1,203	0.318	8.164	9.215	1.051
America	3,471	0.181	5.148	5.696	0.549
Of which: North America	741	0.385	4.635	5.122	0.487
Of which: Latin America and Caribbean	2,730	0.126	5.287	5.852	0.566
Oceania	672	0.342	5.074	5.916	0.842

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms.

This table replicates the baseline results in Table 1, <u>controlling for regional dummies</u>, using the sample of 68 countries for which regional dummies are available (lost countries compared to baseline: Pakistan, Poland, Singapore, Trinidad and Tobago, and Tanzania).

Appendix Table B9 - Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type, controlling for region dummies

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	R^2 without ethnic dummies	R^2 with ethnic dummies	ΔR^2
Breakdown by Question Category					
A: Perceptions of Life	4,090	0.292	6.672	7.456	0.783
B: Environment	930	0.313	6.743	7.636	0.892
C: Work	2,278	0.299	5.531	6.318	0.788
D: Family	1,243	0.320	6.813	7.663	0.851
E: Politics and Society	8,529	0.296	5.992	6.828	0.836
F: Religion and Morals	2,160	0.397	7.505	8.575	1.069
G: National Identity	977	0.397	6.462	7.617	1.155
Breakdown by Question Type					
Binary questions	4,278	0.310	6.394	7.256	0.862
Binary from unordered response questions	6,628	0.256	4.519	5.264	0.745
Scale questions	9,301	0.356	7.628	8.578	0.949

This table replicates the baseline results in Table 2, <u>controlling for regional dummies</u>, using the sample of 68 countries for which regional dummies are available (lost countries compared to baseline: Pakistan, Poland, Singapore, Trinidad and Tobago, and Tanzania).

Appendix Table B10 – Joint Significance of Religion Dummies in Questions from the World Values / European Values Integrated Surveys, by Region

	# of regres-sions	Share of regressions with jointly significant religion dummies	R^2 without religion dummies	R^2 with religion dummies	ΔR^2
Whole sample	21,545	0.362	2.639	4.023	1.384
Africa	3,412	0.370	2.470	3.513	1.043
Of which: Sub-Saharan Africa	2,901	0.411	2.346	3.540	1.194
Of which: North Africa	511	0.139	3.177	3.359	0.182
Europe	7,947	0.336	3.015	4.402	1.388
Of which: Western and Southern Europe	2,365	0.355	3.458	5.098	1.640
Of which: Eastern and Central Europe	5,582	0.329	2.827	4.108	1.281
Asia	5,765	0.436	2.318	3.872	1.553
Of which: East and Southeast Asia	2,204	0.443	2.185	4.074	1.889
Of which: South Asia	849	0.582	2.774	3.630	0.856
Of which: Southwestern and Central Asia	1,505	0.300	2.183	3.130	0.947
Of which: Middle East	1,207	0.490	2.410	4.597	2.187
America	3,751	0.294	2.361	3.667	1.306
Of which: North America	744	0.437	2.945	4.859	1.914
Of which: Latin America and Caribbean	3,007	0.259	2.217	3.372	1.155
Oceania	670	0.358	3.348	5.417	2.069

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms.

This table replicates the baseline results in Table 1, <u>doing the exercise for religion dummies instead of ethnic dummies</u>, using the sample of 72 countries for which both ethnic and religion dummies are available (lost country compared to baseline: Algeria).

Appendix Table B11 - Joint Significance of Religion Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type

	# of regressions	Share of regressions with jointly significant religion dummies	R^2 without religion dummies	R^2 with religion dummies	ΔR^2
Breakdown by Question Category					
A: Perceptions of Life	4,394	0.359	3.230	4.649	1.419
B: Environment	979	0.303	2.222	3.448	1.226
C: Work	2,417	0.304	2.441	3.529	1.088
D: Family	1,343	0.412	3.250	4.529	1.279
E: Politics and Society	9,067	0.327	2.399	3.666	1.266
F: Religion and Morals	2,329	0.562	2.849	5.078	2.229
G: National Identity	1,016	0.353	1.797	3.146	1.348
Breakdown by Question Type					
Binary questions	4,564	0.368	2.812	4.403	1.592
Binary from unordered response questions	7,036	0.300	1.601	2.766	1.165
Scale questions	9,945	0.402	3.294	4.738	1.444

This table replicates the baseline results in Table 2, doing the exercise for religion dummies instead of ethnic dummies, using the sample of 72 countries for which both ethnic and religion dummies are available (lost country compared to baseline: Algeria).

Appendix Table B12 – Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Region, controlling for religion dummies

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	R^2 without ethnic dummies	R^2 with ethnic dummies	ΔR^2
Whole sample	21,107	0.339	4.056	5.248	1.192
Africa	3,311	0.545	3.520	5.085	1.565
Of which: Sub-Saharan Africa	2,719	0.580	3.527	5.293	1.766
Of which: North Africa	592	0.383	3.489	4.132	0.643
Europe	7,745	0.232	4.599	5.457	0.858
Of which: Western and Southern Europe	2,360	0.205	5.323	6.232	0.910
Of which: Eastern and Central Europe	5,385	0.244	4.282	5.118	0.836
Asia	5,642	0.462	3.674	5.371	1.697
Of which: East and Southeast Asia	2,085	0.510	3.794	5.794	2.000
Of which: South Asia	848	0.658	3.976	7.299	3.324
Of which: Southwestern and Central Asia	1,507	0.373	3.114	4.007	0.893
Of which: Middle East	1,202	0.353	3.954	4.987	1.033
America	3,740	0.203	3.808	4.637	0.829
Of which: North America	739	0.440	4.964	5.792	0.828
Of which: Latin America and Caribbean	3,001	0.145	3.523	4.352	0.829
Oceania	669	0.280	5.036	6.006	0.970

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms.

This table replicates the baseline results in Table 1, controlling for religion dummies, using the sample of 72 countries for which religion dummies are available (lost country compared to baseline: Algeria).

Appendix Table B13 - Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type, controlling for religion dummies

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	R^2 without ethnic dummies	R^2 with ethnic dummies	ΔR^2
Breakdown by Question Category					
A: Perceptions of Life	4,292	0.325	4.632	5.803	1.171
B: Environment	968	0.340	3.483	4.832	1.350
C: Work	2,380	0.309	3.564	4.561	0.998
D: Family	1,296	0.333	4.602	5.772	1.170
E: Politics and Society	8,933	0.335	3.710	4.897	1.187
F: Religion and Morals	2,223	0.386	5.164	6.392	1.229
G: National Identity	1,015	0.402	3.246	4.820	1.574
Breakdown by Question Type					
Binary questions	4,413	0.319	4.266	5.394	1.128
Binary from unordered response questions	6,943	0.286	2.788	3.775	0.986
Scale questions	9,751	0.385	4.864	6.231	1.367

This table replicates the baseline results in Table 1, controlling for religion dummies, using the sample of 72 countries for which religion dummies are available (lost country compared to baseline: Algeria).

Appendix Table B14 – Joint Significance of Ethnolinguistic Dummies in Binary Questions from the World Values / European Values Integrated Surveys, by Region, Probit Estimates

	# of regres-sions	Share of regressions w/ jointly significant ethnic dummies	Pseudo-R^2 without ethnic dummies	Pseudo-R^2 with ethnic dummies	ΔR^2
Whole sample	11,304	0.372	2.409	3.622	1.213
Africa	1,899	0.471	2.480	3.891	1.412
Of which: Sub-Saharan Africa	1,424	0.527	2.348	3.995	1.647
Of which: North Africa	475	0.301	2.874	3.580	0.706
Europe	3,981	0.331	2.521	3.511	0.990
Of which: Western and Southern Europe	1,117	0.233	2.975	3.620	0.646
Of which: Eastern and Central Europe	2,864	0.369	2.344	3.468	1.124
Asia	3,005	0.503	2.213	4.077	1.864
Of which: East and Southeast Asia	1,099	0.541	2.239	4.309	2.071
Of which: South Asia	456	0.632	2.673	5.759	3.085
Of which: Southwestern and Central Asia	832	0.416	2.138	3.104	0.967
Of which: Middle East	618	0.458	1.927	3.731	1.804
America	2,057	0.188	2.337	2.949	0.612
Of which: North America	430	0.435	2.677	3.423	0.745
Of which: Latin America and Caribbean	1,627	0.123	2.247	2.824	0.577
Oceania	362	0.262	2.837	3.479	0.642

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms.

This table reproduces Table 1, using <u>only binary questions</u> (or binary questions constructed from multinomials). The estimator is now <u>probit</u>. The joint test on ethnic dummies is now a Wald test. The goodness of fit is the <u>pseudo R-squared</u> (McFadden's).

Appendix Table B15 - Joint Significance of Ethnolinguistic Dummies in Binary Questions from the WVS / EVS Integrated Surveys, by Question Category and Question Type, Probit Estimates

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	Pseudo-R^2 without ethnic dummies	Pseudo-R^2 with ethnic dummies	ΔR^2
Breakdown by Question Category					
A: Perceptions of Life	2,395	0.396	2.809	4.186	1.377
B: Environment	532	0.374	2.230	3.366	1.136
C: Work	1,915	0.359	1.903	2.898	0.995
D: Family	551	0.403	3.117	4.311	1.194
E: Politics and Society	4,474	0.335	2.185	3.208	1.024
F: Religion and Morals	976	0.466	3.439	5.442	2.002
G: National Identity	461	0.423	1.780	3.331	1.551
Breakdown by Question Type					
Binary questions	4,421	0.415	3.316	4.766	1.450
Binary from unordered response questions	6,883	0.344	1.826	2.887	1.061

Note: This result does not change if you break it down by continent: there is little difference in shares of questions with significant ethnolinguistic dummies when the breakdown by category is done continent by continent. R^2 is expressed in % terms.

This table reproduces Table 2, using only binary questions (or binary questions constructed from multinomials). The estimator is now probit. The joint test on ethnic dummies is now a Wald test. The goodness of fit is the pseudo R-squared (McFadden's).

Appendix Table B16 – Joint Significance of Urban Dummies in Questions from the World Values / European Values Integrated Surveys, by Region

	# of regres-sions	Share of regressions with jointly significant urban dummies	R^2 without urban dummies	R^2 with urban dummies	ΔR^2
Whole sample	17,964	0.570	2.710	4.354	1.645
Africa	2,763	0.620	2.696	4.926	2.230
Of which: Sub-Saharan Africa	1,857	0.621	2.648	5.153	2.505
Of which: North Africa	906	0.618	2.793	4.461	1.668
Europe	6,769	0.556	3.022	4.331	1.309
Of which: Western and Southern Europe	1,886	0.301	3.538	4.317	0.779
Of which: Eastern and Central Europe	4,883	0.655	2.823	4.336	1.513
Asia	4,512	0.706	2.431	4.731	2.299
Of which: East and Southeast Asia	1,991	0.684	2.113	4.131	2.018
Of which: South Asia	852	0.894	2.693	4.726	2.033
Of which: Southwestern and Central Asia	564	0.768	2.250	4.898	2.647
Of which: Middle East	1,105	0.569	2.895	5.730	2.834
America	3,515	0.426	2.388	3.506	1.118
Of which: North America	746	0.374	3.098	3.761	0.663
Of which: Latin America and Caribbean	2,769	0.440	2.197	3.438	1.241
Oceania	405	0.170	3.470	4.012	0.543

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms. Maximal number of urban categories in each country: 8, defined as follows: towns under 2,000, 2,000 to 5,000, 5,000 to 10,000, 10,000 to 20,000, 20,000 to 50,000, 50,000 to 100,000, 100,000 to 500,000, 500,000 and more.

This table replicates the baseline results in Table 1, <u>doing the exercise for urban dummies instead of ethnic dummies</u>, using the sample of 60 countries for which both ethnic and urban dummies are available. Urban dummies are created from variable x049 in the WVS-EVS integrated file (size of the city where the interview was conducted). Dummies were created for each answer category, and dummies were jointly entered in our baseline regression for each question, instead of ethnic dummies.

Appendix Table B17 - Joint Significance of Urban Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type

	# of regres-sions	Share of regressions with jointly significant urban dummies	R^2 without urban dummies	R^2 with urban dummies	ΔR^2
Breakdown by Question Category					
A: Perceptions of Life	3,702	0.545	3.262	4.825	1.564
B: Environment	780	0.576	2.184	4.222	2.038
C: Work	1,964	0.563	2.393	3.759	1.367
D: Family	1,135	0.570	3.302	4.898	1.596
E: Politics and Society	7,518	0.556	2.423	4.083	1.660
F: Religion and Morals	2,030	0.657	3.308	5.145	1.837
G: National Identity	835	0.601	1.824	3.576	1.752
Breakdown by Question Type					
Binary questions	3,816	0.550	2.854	4.476	1.622
Binary from unordered response questions	5,832	0.510	1.601	2.923	1.322
Scale questions	8,316	0.620	3.421	5.302	1.881

Note: Maximal number of urban categories in each country: 8, defined as follows: towns under 2,000, 2,000 to 5,000, 5,000 to 10,000, 10,000 to 20,000, 20,000 to 50,000, 50,000 to 100,000, 100,000 to 500,000, 500,000 and more.

This table replicates the baseline results in Table 2, <u>doing the exercise for urban dummies instead of ethnic dummies</u>, using the sample of 60 countries for which both ethnic and urban dummies are available. Urban dummies are created from variable x049 in the WVS-EVS integrated file (size of the city where the interview was conducted). Dummies were created for each answer category, and dummies were jointly entered in our baseline regression for each question, instead of ethnic dummies.

Appendix Table B18 – Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Region, controlling for urban dummies

	# of regres-sions	Share of regressions w/ jointly significant ethnic dummies	R2 without ethnic dummies	R2 with ethnic dummies	∆R2
Whole sample	17,714	0.400	4.444	5.803	1.360
Africa	2,853	0.533	4.731	6.727	1.996
Of which: Sub-Saharan Africa	1,954	0.634	4.891	7.530	2.639
Of which: North Africa	899	0.315	4.385	4.982	0.597
Europe	6,560	0.346	4.566	5.559	0.993
Of which: Western and Southern Europe	1,880	0.273	4.583	5.374	0.791
Of which: Eastern and Central Europe	4,680	0.375	4.559	5.633	1.075
Asia	4,392	0.533	4.786	6.869	2.083
Of which: East and Southeast Asia	1,871	0.594	4.069	6.233	2.164
Of which: South Asia	852	0.663	5.047	8.276	3.229
Of which: Southwestern and Central Asia	564	0.342	4.901	5.640	0.739
Of which: Middle East	1,105	0.428	5.741	7.491	1.749
America	3,504	0.220	3.601	4.285	0.684
Of which: North America	741	0.516	3.847	4.717	0.870
Of which: Latin America and Caribbean	2,763	0.140	3.535	4.169	0.634
Oceania	405	0.432	4.014	4.826	0.812

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms. Maximal number of urban categories in each country: 8, defined as follows: towns under 2,000, 2,000 to 5,000, 5,000 to 10,000, 10,000 to 20,000, 20,000 to 50,000, 50,000 to 100,000, 100,000 to 500,000, 500,000 and more.

This table reproduces Table 1, adding urban dummies to the baseline specification, using the sample of 60 countries for which both ethnic and urban dummies are available. Urban dummies are created from variable x049 in the WVS-EVS integrated file (size of the city where the interview was conducted). Dummies were created for each answer category, and dummies were jointly entered in our baseline regression for each question, in addition to ethnic dummies. Maximal number of urban categories in each country: 8, defined as follows: towns under 2,000, 2,000 to 5,000, 5,000 to 10,000, 10,000 to 20,000, 20,000 to 50,000, 50,000 to 100,000, 100,000 to 500,000, 500,000 and more.

Appendix Table B19 - Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type, controlling for urban dummies

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	R2 without ethnic dummies	R2 with ethnic dummies	ΔR2
Breakdown by Question Category					
A: Perceptions of Life	3,628	0.392	4.927	6.242	1.315
B: Environment	781	0.392	4.270	5.705	1.436
C: Work	1,931	0.374	3.925	5.000	1.076
D: Family	1,090	0.406	4.949	6.310	1.362
E: Politics and Society	7,510	0.377	4.157	5.462	1.306
F: Religion and Morals	1,935	0.497	5.289	7.007	1.719
G: National Identity	839	0.461	3.673	5.461	1.788
Breakdown by Question Type					
Binary questions	3,672	0.400	4.562	5.896	1.335
Binary from unordered response questions	5,812	0.333	3.004	4.083	1.079
Scale questions	8,230	0.446	5.407	6.976	1.569

Note: Maximal number of urban categories in each country: 8, defined as follows: towns under 2,000, 2,000 to 5,000, 5,000 to 10,000, 10,000 to 20,000, 20,000 to 50,000, 50,000 to 100,000, 100,000 to 500,000, 500,000 and more.

This table reproduces Table 2, adding urban dummies to the baseline specification, using the sample of 60 countries for which both ethnic and urban dummies are available. Urban dummies are created from variable x049 in the WVS-EVS integrated file (size of the city where the interview was conducted). Dummies were created for each answer category, and dummies were jointly entered in our baseline regression for each question, in addition to ethnic dummies.

Appendix Table B20 – Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Region, not controlling for income and education but only for age and sex.

	# of regres-sions	Share of regressions w/ jointly significant ethnic dummies	R2 without ethnic dummies	R2 with ethnic dummies	∆R2
Whole sample	21,467	0.455	1.349	2.844	1.495
Africa	3,623	0.580	1.011	2.811	1.801
Of which: Sub-Saharan Africa	2,724	0.645	0.937	3.093	2.156
Of which: North Africa	899	0.384	1.233	1.958	0.724
Europe	7,769	0.386	1.735	2.891	1.155
Of which: Western and Southern Europe	2,369	0.323	1.864	2.748	0.884
Of which: Eastern and Central Europe	5,400	0.413	1.679	2.954	1.275
Asia	5,654	0.597	1.132	3.428	2.296
Of which: East and Southeast Asia	2,088	0.659	0.956	3.548	2.591
Of which: South Asia	852	0.689	1.213	4.925	3.712
Of which: Southwestern and Central Asia	1,511	0.506	1.100	2.395	1.295
Of which: Middle East	1,203	0.538	1.419	3.458	2.039
America	3,749	0.287	1.110	1.919	0.809
Of which: North America	741	0.564	1.561	2.578	1.017
Of which: Latin America and Caribbean	3,008	0.219	0.998	1.756	0.758
Oceania	672	0.335	1.871	2.721	0.850

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms.

This table reproduces Table 1, <u>excluding income and education controls from the regressions</u>, but keeping the age and sex controls. We maintained the same sample as in Table 1 for comparability.

Appendix Table B21 - Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type, not controlling for income and education but only for age and sex

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	R2 without ethnic dummies	R2 with ethnic dummies	ΔR2
Breakdown by Question Category					
A: Perceptions of Life	4,380	0.451	1.812	3.237	1.425
B: Environment	971	0.450	0.791	2.397	1.607
C: Work	2,409	0.431	1.015	2.271	1.257
D: Family	1,319	0.475	2.188	3.669	1.481
E: Politics and Society	9,046	0.433	1.027	2.442	1.415
F: Religion and Morals	2,316	0.540	2.126	4.059	1.933
G: National Identity	1,026	0.514	0.694	2.674	1.979
Breakdown by Question Type					
Binary questions	4,550	0.451	1.389	2.915	1.527
Binary from unordered response questions	7,029	0.386	0.836	1.995	1.159
Scale questions	9,888	0.507	1.696	3.414	1.718

This table reproduces Table 2, excluding income and education controls from the regressions, but keeping the age and sex controls. We maintained the same sample as in Table 2 for comparability.

Appendix Table B22 – Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Region, controlling for income, education, age, sex marital status, whether respondent has children, and respondent occupation

	# of regres-sions	Share of regressions w/ jointly significant ethnic dummies	R2 without ethnic dummies	R2 with ethnic dummies	ΔR2
Whole sample	20,520	0.387	4.932	6.390	1.458
Africa	3,622	0.439	5.582	7.549	1.967
Of which: Sub-Saharan Africa	2,723	0.475	5.826	8.198	2.372
Of which: North Africa	899	0.330	4.842	5.583	0.740
Europe	7,682	0.336	5.112	6.181	1.069
Of which: Western and Southern Europe	2,283	0.235	5.633	6.371	0.738
Of which: Eastern and Central Europe	5,399	0.378	4.892	6.100	1.209
Asia	5,331	0.535	4.243	6.367	2.124
Of which: East and Southeast Asia	2,088	0.580	4.090	6.395	2.305
Of which: South Asia	852	0.662	4.945	8.524	3.579
Of which: Southwestern and Central Asia	1,511	0.458	3.967	5.058	1.090
Of which: Middle East	880	0.441	4.399	6.460	2.061
America	3,213	0.212	4.813	5.652	0.839
Of which: North America	741	0.480	4.451	5.356	0.905
Of which: Latin America and Caribbean	2,472	0.131	4.922	5.741	0.819
Oceania	672	0.351	5.394	6.233	0.839

Note: North America is defined here as Canada and the US. Mexico is included with Latin America and the Caribbean. R^2 is expressed in % terms.

This table reproduces Table 1, <u>adding controls</u> for marital status (a dummy for married / unmarried, based on WVS-EVS variable x007), whether the respondent has children (a dummy=1 if the respondent has children, based on WVS-EVS variable x011) and a series of categorical variables for various occupations (based on WVS-EVS variable x036: profession of respondent). We lose 4 countries compared to the baseline: Guatemala, Iran, Israel and Peru, for which data on the additional controls is not available.

Appendix Table B23 - Joint Significance of Ethnolinguistic Dummies in Questions from the World Values / European Values Integrated Surveys, by Question Category and Question Type, controlling for income, education, age, sex marital status, whether respondent has children, and respondent occupation

	# of regres-sions	Share of regressions with jointly significant ethnic dummies	R2 without ethnic dummies	R2 with ethnic dummies	ΔR2
Breakdown by Question Category					
A: Perceptions of Life	4,168	0.382	5.527	6.946	1.419
B: Environment	933	0.387	4.575	6.087	1.512
C: Work	2,316	0.370	4.463	5.642	1.179
D: Family	1,264	0.382	5.563	6.940	1.377
E: Politics and Society	8,628	0.365	4.678	6.090	1.412
F: Religion and Morals	2,221	0.472	5.490	7.299	1.809
G: National Identity	990	0.457	4.003	5.943	1.941
Breakdown by Question Type					
Binary questions	4,360	0.386	5.044	6.485	1.440
Binary from unordered response questions	6,745	0.327	3.672	4.842	1.170
Scale questions	9,415	0.430	5.782	7.454	1.672

This table reproduces Table 2, adding controls for marital status (a dummy for married / unmarried, based on WVS-EVS variable x007), whether the respondent has children (a dummy=1 if the respondent has children, based on WVS-EVS variable x011) and a series of categorical variables for various occupations (based on WVS-EVS variable x036: profession of respondent). We lose 4 countries compared to the baseline: Guatemala, Iran, Israel and Peru, for which data on the additional controls is not available.

Appendix B4: Robustness Tables for Section 4 Results

Table B24: Additional Results Corresponding to Table 3 in the main text

Panel A – Correlations among Cultural Diversity and χ^2 by Question Category

	CF Overall	CF A	CF B	CF C	CF D	CF E	CF F
CF Category A	0.432	1					
CF Category B	0.386	0.140	1				
CF Category C	0.732	0.185	-0.040	1			
CF Category D	0.596	0.122	0.372	0.283	1		
CF Category E	0.757	0.302	0.277	0.565	0.477	1	
CF Category F	0.856	0.077	0.254	0.618	0.492	0.482	1
CF Category G	-0.009	-0.132	0.234	-0.354	0.411	-0.104	0.074

(Based on 76 observations)

	χ^2 Overall	χ^2 A	χ^2 B	χ^2 C	χ^2 D	χ^2 E	χ^2 F
$\chi2$ Category A	0.967	1					
$\chi2$ Category B	0.765	0.716	1				
$\chi2$ Category C	0.798	0.738	0.923	1			
$\chi2$ Category D	0.921	0.915	0.685	0.660	1		
$\chi2$ Category E	0.978	0.918	0.776	0.808	0.894	1	
$\chi2$ Category F	0.941	0.903	0.643	0.681	0.825	0.876	1
$\chi2$ Category G	0.632	0.555	0.449	0.420	0.664	0.634	0.554

(Based on 76 observations)

Question categories are defined as follows: A: Perceptions of Life (42 questions); B: Environment (4 questions); C: Work (25 questions); D: Family (12 questions); E: Politics and Society (59 questions); F: Religion and Morale (30 questions); G: National Identity (3 questions).

Panel B - Correlations among Cultural Diversity and χ^2 by Question Type

	CF All	CF Binary	CF Scale
CF Binary	0.800	1	
CF Scale	0.918	0.667	1
CF Unordered	0.603	0.588	0.507

	χ^2 All	χ^2 Binary	χ^2 Scale
χ^2 Binary	0.904	1	
χ^2 Scale	0.996	0.880	1
χ^2 Unordered	0.954	0.885	0.937

Binary: 49 questions; unordered response questions: 26 questions; scale: 100 questions

Panel C: Correlations between Diversity Measures by Question Category

	Correlations with		
	Cultural Fractionalization	ELF	F_{ST}
Question Category A			
ELF	0.296	1	
F_{ST}	0.465	0.564	1
χ^2	0.457	0.577	0.986
Question Category B			
ELF	-0.118	1	
F_{ST}	-0.279	0.556	1
χ^2	-0.228	0.561	0.989
Question Category C			
ELF	-0.033	1	
F_{ST}	0.133	0.623	1
χ^2	0.302	0.556	0.890
Question Category D			
ELF	-0.072	1	
F_{ST}	0.073	0.630	1
χ^2	0.086	0.643	0.977
Question Category E			
ELF	0.165	1	
F_{ST}	0.198	0.620	1
χ^2	0.231	0.637	0.983
Question Category F			
ELF	-0.255	1	
F_{ST}	-0.014	0.477	1
χ^2	0.119	0.476	0.892
Question Category G			
ELF	-0.221	1	
F_{ST}	-0.013	0.430	1
χ^2	0.137	0.359	0.913

(Based on 76 observations)
Question categories are defined as follows: A: Perceptions of Life (42 questions); B: Environment (4 questions); C: Work (25 questions); D: Family (12 questions); E: Politics and Society (59 questions); F: Religion and Morale (30 questions); G: National Identity (3 questions).

Panel D: Correlations between Diversity Measures by Question Type

	Cultural Fractionalization	ELF	F_{ST}
Binary Response Questions (49 questions)			
ELF	-0.075	1	
F_{ST}	0.157	0.543	1
χ^2	0.171	0.540	1
Unordered Response Questions (26 questions)			
ELF	-0.282	1	
F_{ST}	-0.022	0.542	1
χ^2	-0.036	0.567	0.982
Scale Response Questions (100 questions)			
ELF	0.111	1	
F_{ST}	0.267	0.631	1
χ^2	0.287	0.626	0.986

Panel E: Correlations between Diversity Measures for Restricted Set of 9 Questions

	Cultural Fractionalization	ELF	F_{ST}
Restricted Set of 9 Questions Used in Literature			
ELF	-0.222	1	
F_{ST}	0.022	0.551	1
χ^2	-0.023	0.586	0.969

Panel F: Correlation of Indices

	Cultural Fractionalization	ELF
FACTOR 1		
ELF	0.034	
Chi-square	0.057	0.554**
FACTOR 2		
ELF	-0.333*	
Chi-square	0.328*	0.365*
FACTOR 3		
ELF	0.155	
Chi-square	0.029	0.622**
FACTOR 4		
ELF	-0.059	
Chi-square	-0.026	0.698**
FACTOR 5		
ELF	0.217	
Chi-square	0.257	0.536**
AVERAGE Of 5 FACTORS		
ELF	-0.044	
Chi-square	0.118	0.690**

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

Table B25 – Summary Statistics for the Main Indices of Regional Heterogeneity, Cultural Diversity, F_{ST} and χ^2

The panels below replicate our baseline analysis of diversity measures taking first-level administrative regions as the basis for group identity. We recomputed ELF (the probability that two randomly drawn individuals belong to different groups as defined above), F_{ST} and Chi-square.

Panel A: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Cultural Fractionalization	0.494	0.036	0.405	0.599
ELF	0.834	0.157	0.000	0.973
F_{ST}	0.031	0.022	0.000	0.134
χ^2	0.072	0.049	0.000	0.262

(Summary statistics based on 89 observations)

Panel B: Correlations

	Cultural Fractionalization	ELF	F_{ST}	χ^2
Cultural Fractionalization	1			
ELF	0.022	1		
F_{ST}	-0.008	0.348**	1	
χ^2	0.052	0.361**	0.985**	1

(** Significant at the 1% level; * Significant at 5%; correlations based on 89 observations)

Appendix Table B26 – Summary Statistics for the Main Indices of Large-Small Town Heterogeneity, Cultural Diversity, F_{ST} and χ^2

The panels below replicate our baseline analysis of diversity measures taking town size as the basis for group identity. In this version we consider two groups: those living in towns smaller than 500,000 inhabitants, and those living in towns bigger than this threshold. We recomputed ELF (the probability that two randomly drawn individuals belong to different groups as defined above), F_{ST} and Chi-square.

Panel A: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Cultural Fractionalization	0.490	0.036	0.404	0.606
ELF	0.247	0.163	0.000	0.500
F_{ST}	0.003	0.003	0.000	0.013
χ^2	0.006	0.005	0.000	0.025

(Summary statistics based on 82 observations)

Panel B: Correlations

	Cultural Fractionalization	ELF	F_{ST}	χ^2
Cultural Fractionalization	1			
ELF	-0.218*	1		
F_{ST}	-0.278*	0.530**	1	
χ^2	-0.233*	0.533**	0.984**	1

(** Significant at the 1% level; * Significant at 5%; correlations based on 82 observations)

Appendix Table B27 – Summary Statistics for the Main Indices of Different Sized Town Heterogeneity, Cultural Diversity, FST and χ^2

The panels below replicate our baseline analysis taking town size as the basis for group identity. In this version we consider eight groups: inhabitants of towns < 2,000; 2,000-5,000; 5,000-10,000; 10,000-20,000; 20,000-50,000; 50,000-100,000; 100,000-500,000; > 500,000. We recomputed ELF (the probability that two randomly drawn individuals belong to different groups as defined above), F_{ST} and Chi-square.

Panel A: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Cultural Fractionalization	0.490	0.036	0.404	0.606
ELF	0.761	0.143	0.000	0.871
F_{ST}	0.013	0.010	0.000	0.081
χ^2	0.029	0.019	0.000	0.101

(Summary statistics based on 82 observations)

Panel B: Correlations

	Cultural Fractionalization	ELF	F_{ST}	χ^2
Cultural Fractionalization	1			
ELF	-0.357**	1		
F_{ST}	-0.266**	0.086	1	
χ^2	-0.104	0.052	0.913**	1

(** Significant at the 1% level; correlations based on 82 observations)

Appendix Table B28 – Summary Statistics for Cultural Diversity and F_{ST} that Incorporate Distances between Answers

Panel A: Summary Statistics

	CF (Distance)	F_{ST} (Distance)
Mean	0.333	0.013
Standard Deviation	0.023	0.012

(Summary statistics based on 76 observations)

Panel B: Correlations

	CF (Distance)	ELF
ELF	0.049	1
F_{ST} (Distance)	0.123	0.628**

(** Significant at the 1% level; correlations based on 76 observations)

Panel C: Summary Statistics, Only Scale Questions

	CF (Distance)	F_{ST} (Distance)
Mean	0.265	0.013
Standard Deviation	0.021	0.013

(Summary statistics based on 76 observations)

Panel D: Correlations, Only Scale Questions

	CF (Distance)	ELF
ELF	0.228*	1
F_{ST} (Distance)	0.200	0.636**

(** Significant at the 1% level; * Significant at the 5% level; correlations based on 76 observations)

Appendix Table B29 – Summary Statistics for the Main Indices of Ethnic Heterogeneity, Cultural Diversity, F_{ST} and χ^2 for Afrobarometer (5th wave, 2014)

Panel A: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Cultural Fractionalization	0.613	0.020	0.560	0.653
ELF	0.796	0.132	0.292	0.961
F_{ST}	0.034	0.016	0.005	0.080
χ^2	0.120	0.061	0.017	0.304

(Summary statistics based on 28 observations)

Panel B: Correlations

	Cultural Fractionalization	ELF	F_{ST}	χ^2
Cultural Fractionalization	1			
ELF	0.291	1		
F_{ST}	0.442*	0.590**	1	
χ^2	0.394*	0.599**	0.985**	1

(** Significant at the 1% level; * Significant at 5%; correlations based on 28 observations)

Appendix Table B30 – Summary Statistics for the Main Indices of Ethnic Heterogeneity, Cultural Diversity, F_{ST} and χ^2 for Latinobarómetro (2007, 2008, 2009 and 2010 waves)

Panel A: Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Cultural Fractionalization	0.527	0.010	0.509	0.546
ELF	0.547	0.137	0.294	0.762
F_{ST}	0.009	0.002	0.007	0.014
χ^2	0.027	0.006	0.019	0.039

(Summary statistics based on 18 observations)

Panel B: Correlations

	Cultural Fractionalization	ELF	F_{ST}	χ^2
Cultural Fractionalization	1			
ELF	0.729**	1		
F_{ST}	0.453	0.371	1	
χ^2	0.413	0.418	0.956**	1

(** Significant at the 1% level; * Significant at 5%; correlations based on 18 observations)

Appendix Table B31 – Summary Statistics for the Main Indices of Ethnic Heterogeneity, Cultural Diversity, F_{ST} and χ^2: Continents and Regions

	CF	ELF	F_{ST}	χ^2
Whole Sample	**0.529**	**0.390**	**0.012**	**0.029**
Africa	**0.515**	**0.583**	**0.014**	**0.038**
Of which: Sub-Saharan Africa	0.536	0.653	0.017	0.046
Of which: North Africa	0.454	0.374	0.005	0.014
Europe	**0.543**	**0.238**	**0.008**	**0.020**
Of which: Western and Southern Europe	0.559	0.122	0.005	0.016
Of which: Eastern and Central Europe	0.535	0.296	0.010	0.023
Asia	**0.510**	**0.421**	**0.018**	**0.042**
Of which: East and Southeast Asia	0.521	0.415	0.019	0.045
Of which: South Asia	0.482	0.586	0.027	0.062
Of which: Southwestern and Central Asia	0.516	0.271	0.011	0.027
Of which: Middle East	0.502	0.485	0.016	0.036
America	**0.541**	**0.458**	**0.006**	**0.016**
Of which: North America	0.555	0.481	0.009	0.021
Of which: Latin America and Caribbean	0.538	0.454	0.005	0.015
Oceania	**0.545**	**0.142**	**0.006**	**0.019**

Appendix Table B32 – Correlations of Different Variables with CF and χ^2 by Question Type

	Cultural Fractionalization	χ^2
Binary Questions		
Linguistic Diversity (Alesina et al., 2003)	-0.034	0.486**
Ethnic Diversity (Alesina et al., 2003)	-0.133	0.268*
Religious Diversity (Alesina et al., 2003)	0.319**	0.123
Genetic Diversity (Ashraf & Galor, 2013)	-0.239*	0.205
RQ Ethnolinguistic Polarization (EMR, 2012)	-0.120	0.189
ER Ethnolinguistic Polarization (EMR, 2012)	0.036	0.250*
Ethnic Diversity (Fearon, 2003)	-0.085	0.380**
Ethnic Greenberg (Fearon, 2003)	-0.041	0.457**
Ethnic Inequality GREG	-0.168	0.037
Ethnic Inequality Ethnologue	-0.206	0.040
Unordered Response Questions		
Linguistic Diversity (Alesina et al., 2003)	-0.254*	0.539**
Ethnic Diversity (Alesina et al., 2003)	-0.331**	0.329**
Religious Diversity (Alesina et al., 2003)	0.163	0.021
Genetic Diversity (Ashraf & Galor, 2013)	-0.178	0.196
RQ Ethnolinguistic Polarization (EMR, 2012)	-0.005	0.225
ER Ethnolinguistic Polarization (EMR, 2012)	0.072	0.254*
Ethnic Diversity (Fearon, 2003)	-0.194	0.384**
Ethnic Greenberg (Fearon, 2003)	-0.090	0.468**
Ethnic Inequality GREG	-0.328**	0.088
Ethnic Inequality Ethnologue	-0.466**	0.070
Scale Questions		
Linguistic Diversity (Alesina et al., 2003)	0.171	0.590**
Ethnic Diversity (Alesina et al., 2003)	0.053	0.395**
Religious Diversity (Alesina et al., 2003)	0.335**	0.086
Genetic Diversity (Ashraf & Galor, 2013)	-0.068	0.183
RQ Ethnolinguistic Polarization (EMR, 2012)	0.019	0.174
ER Ethnolinguistic Polarization (EMR, 2012)	0.088	0.220
Ethnic Diversity (Fearon, 2003)	0.162	0.440**
Ethnic Greenberg (Fearon, 2003)	0.153	0.494**
Ethnic Inequality GREG	-0.107	0.130
Ethnic Inequality Ethnologue	-0.143	0.204

* significant at 5%; ** significant at 1%

Appendix Table B33 – Persistence of CF and χ^2: Correlations across Waves

	Countries with Ethnic Identifiers		All Countries
	CF	χ^2	CF
Wave 3-Wave 4	0.681*	0.970**	0.886**
Observations	(9)	(9)	(20)
Wave 4-Wave 5	0.704*	0.953**	0.907**
Observations	(11)	(11)	(29)
Wave 3-Wave 5	0.883*	0.709*	0.577**
Observations	(11)	(11)	(20)

* significant at 5%; ** significant at 1%

Appendix Figure B1: Cultural Fractionalization (CF)
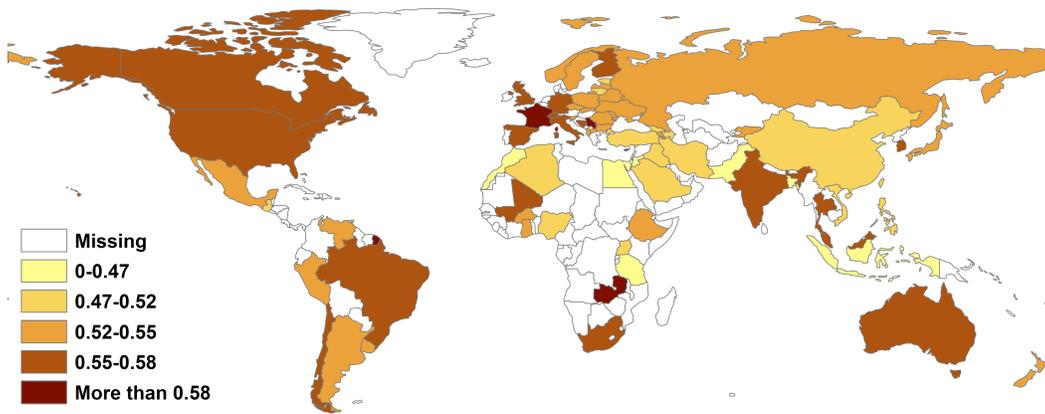
Appendix Figure B2: ELF
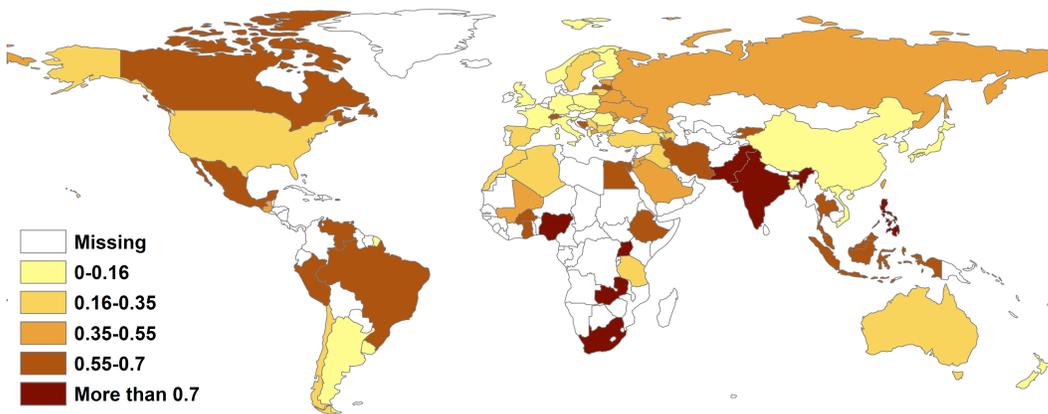
Appendix Figure B3: Chi-Square
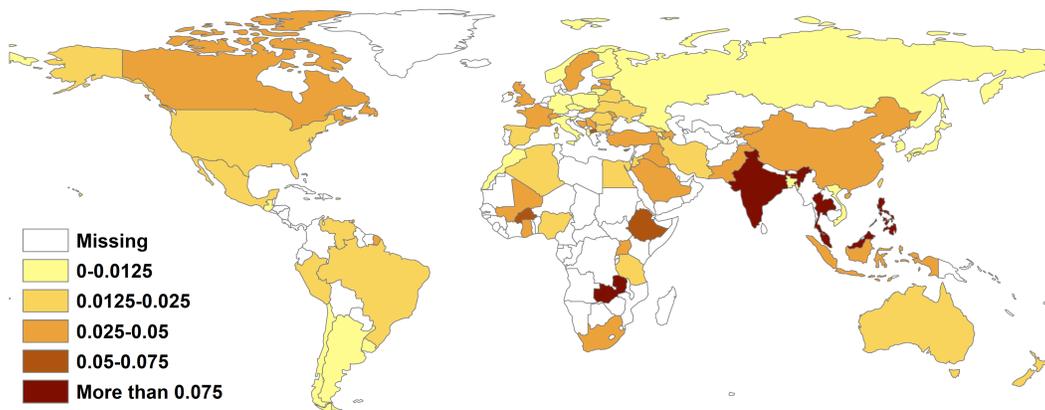
Appendix Figure B4 – CF, χ^2 and ELF: Scatterplots

Panel A: CF vs ELF
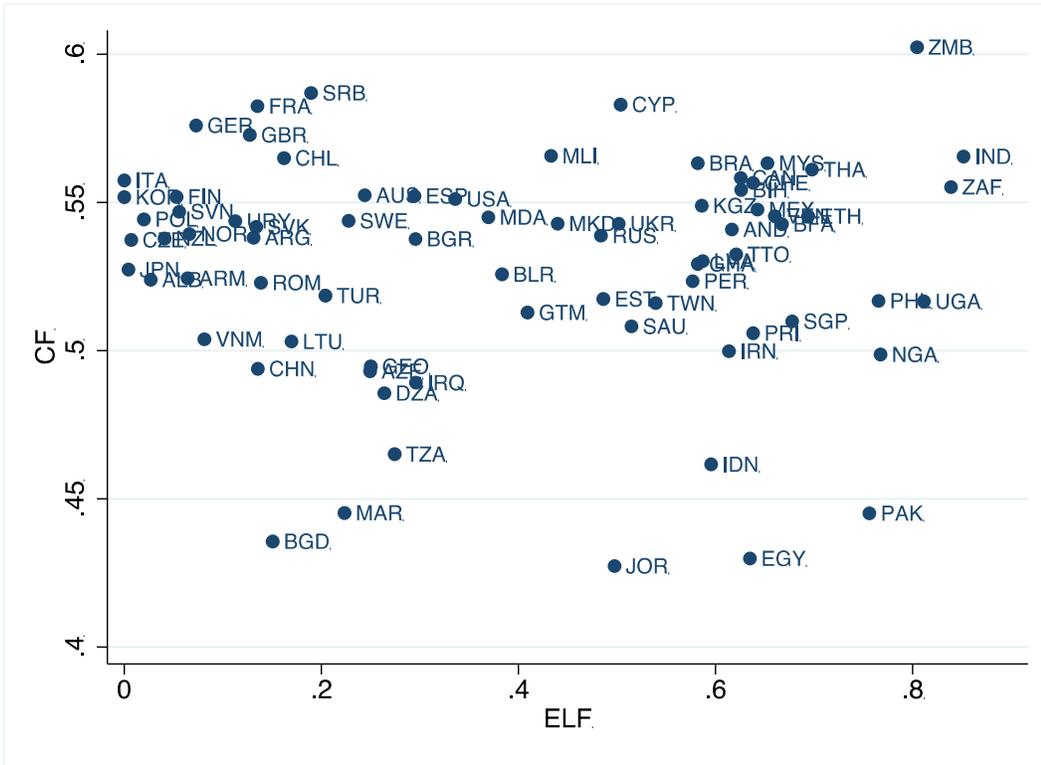
Panel B: χ^2 vs ELF
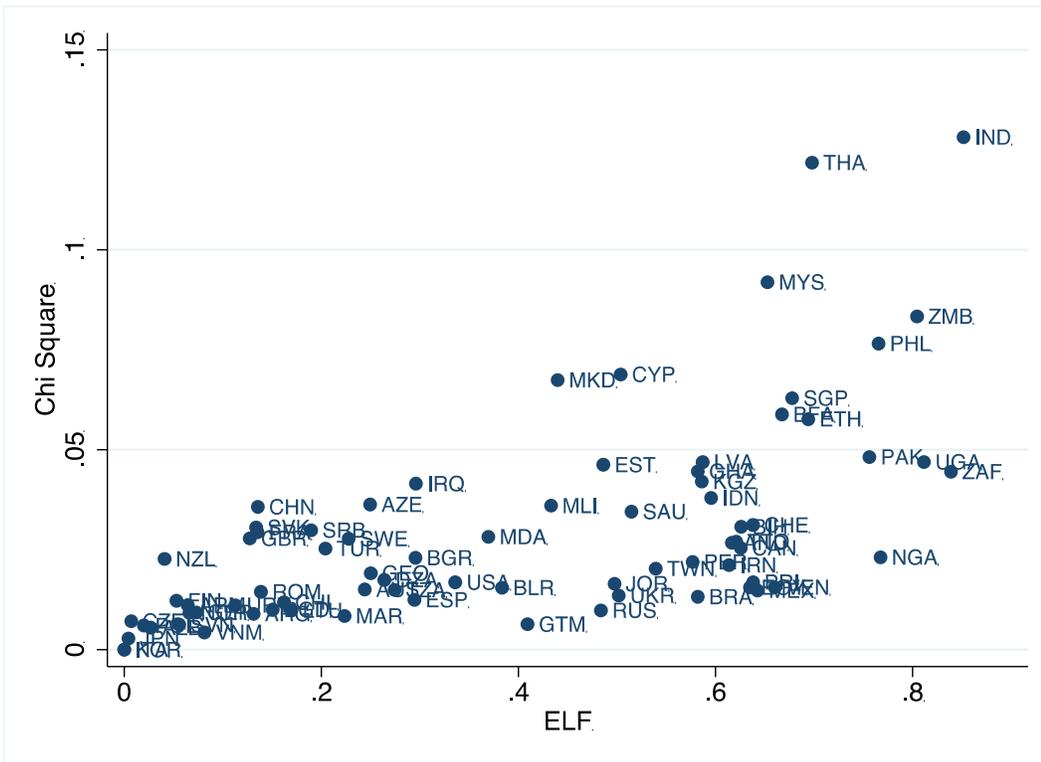
Panel C: χ² vs CF
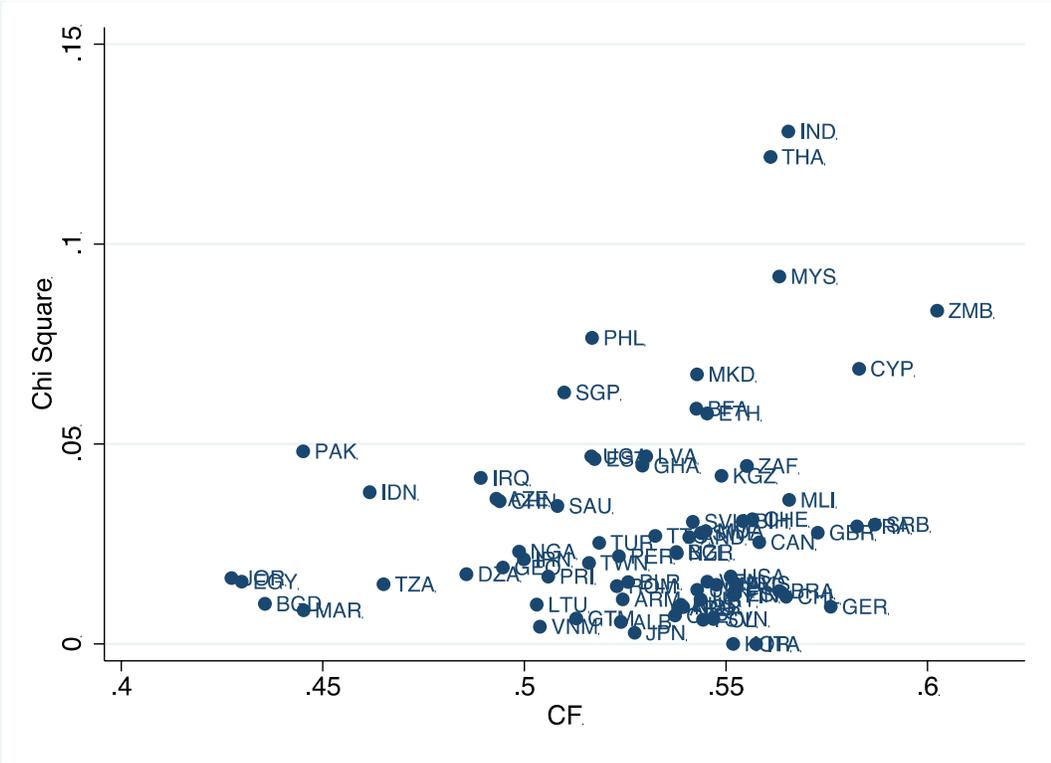
Appendix Figure B5: Genetic Diversity and Cultural Diversity
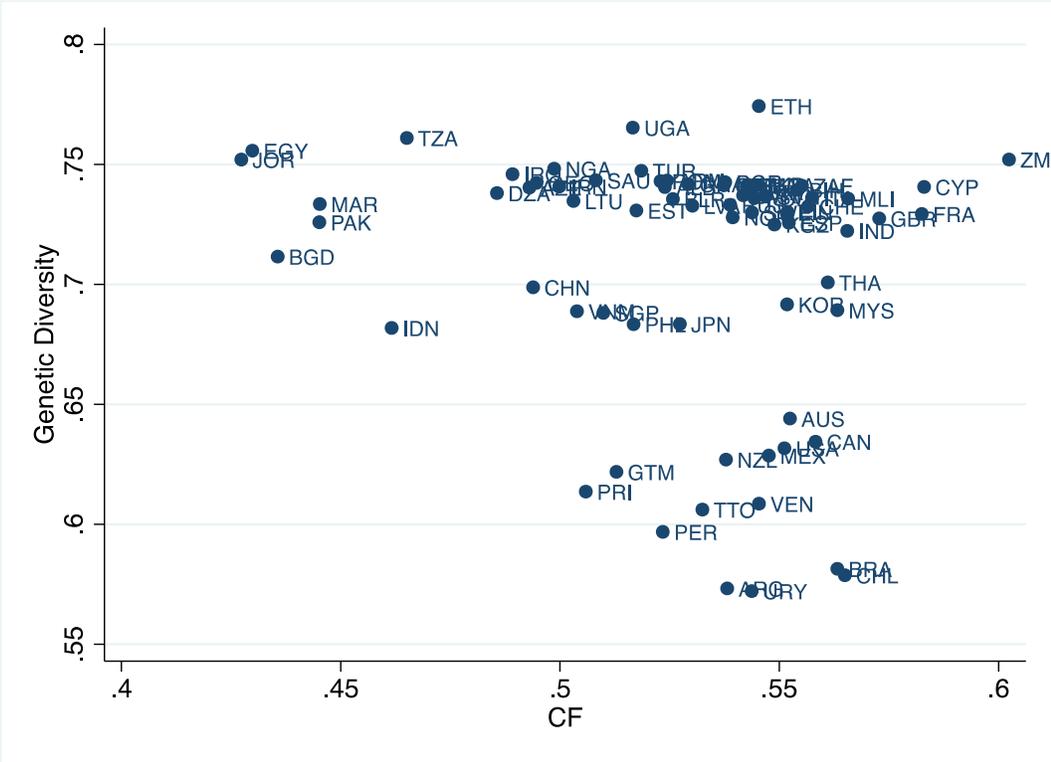
Appendix B5: Robustness Tables for Section 5 - Civil Conflict, Public Goods and Income Regressions

Appendix Table B34 – Endogeneity of Chi-Square and Cultural Fractionalization
(Dependent Variable: Incidence of Civil Conflict)

|	(1) Persistent Questions	(2) Non-Persistent Questions	(3) Post-1970	(4) Post-1970 Cohort < 1950
Cultural Fractionalization	-0.283 [-1.719]	-0.262** [-3.806]	-0.504** [-2.636]	-0.358** [-2.737]
Ethnolinguistic Fractionalization	-0.027 [-1.317]	-0.024 [-1.269]	-0.022 [-0.645]	-0.017 [-0.657]
Chi-square	0.685** [3.329]	0.772** [3.634]	1.071** [2.600]	0.370** [2.650]
Observations	2,971	2,921	1,785	1,636
Pseudo R-squared	0.757	0.759	0.793	0.796

Robust z statistics in brackets; * significant at 5%; ** significant at 1%. Logit estimation, based on at most 70 countries, standard errors clustered at country level. The columns report marginal effects.

All columns include controls for lagged war, log lagged GDP per capita, log lagged population, fraction mountainous terrain, non-contiguous territory dummy, oil dummy, new state dummy, instability dummy, democracy lagged (Polity 2), Latin America and Caribbean dummy, Sub-Saharan Africa dummy, East and Southeast Asia dummy.

Appendix Table B35: Adding Geographic Controls from Michalopoulos (2012) to the baseline conflict regression.

	(1)	(2)
	Baseline Specification	**Adding Michalopoulos Geography Controls**
Cultural Fractionalization	**-0.391****	**-0.524****
	[-3.330]	**[-2.826]**
Ethnolinguistic Fractionalization	**-0.037**	**-0.051**
	[-1.769]	**[-1.746]**
Chi-square	**0.870****	**1.182****
	[3.823]	**[3.176]**
Lagged War	0.840**	0.818**
	[27.877]	[24.872]
Log Lagged GDP per capita	0.003	0.002
	[0.621]	[0.384]
Log Lagged Population	0.012**	0.017**
	[4.183]	[3.677]
% Mountainous Terrain	0.041**	0.059*
	[2.602]	[2.435]
Country with non-connected territories	0.026	0.025
	[1.596]	[1.621]
Oil	0.022	0.025
	[1.618]	[1.762]
New State	0.210*	0.215
	[2.143]	[1.862]
Instability	-0.009	-0.007
	[-1.080]	[-0.823]
Democracy Lagged (Polity 2)	0.003	0.003
	[0.490]	[0.380]
Variation in Elevation		-0.061*
		[-2.151]
Variation in Agricultural Suitability		0.088
		[1.567]
Mean Elevation		0.025
		[1.283]
Mean Agricultural Suitability		-0.009
		[-0.366]
Mean Precipitation		0.072
		[0.768]
Mean Temperature		0.046
		[0.057]
Observations	2,921	2,702
Pseudo R-squared	0.758	0.762

Robust z statistics in brackets; * significant at 5%; ** significant at 1%. Logit estimation, based on at most 70 countries from 1945 to 1999, standard errors clustered at the country level. The columns report marginal effects. Mean precipitation and mean temperature have been multiplied by 1000 for readability.
Both columns include dummies for Latin America and Caribbean, Sub-Saharan Africa, East and Southeast Asia.

Appendix Table B36 - Incidence of Civil Conflict and Diversity, by Question Category and Type
(Dependent Variable: Incidence of Civil Conflict)

Panel A – By Question Category

	A	B	C	D	E	F	G	No G
	(42q)	(4q)	(25q)	(12q)	(59q)	(30q)	(3q)	(172q)
Cultural Fractionalization	-0.200	0.025	-0.062	-0.123	-0.135	-0.119**	-0.035	-0.376**
	[-1.876]	[0.400]	[-1.062]	[-1.111]	[-0.909]	[-3.350]	[-1.384]	[-3.336]
Ethnolinguistic Fractionalization	-0.016	0.021	0.011	-0.034	-0.028	-0.007	0.000	-0.035
	[-0.962]	[1.219]	[0.493]	[-1.603]	[-1.377]	[-0.440]	[-0.011]	[-1.732]
Chi Square	0.756**	-0.009	0.075	0.843**	0.687**	0.394**	0.202*	0.850**
	[3.022]	[-0.072]	[0.596]	[3.292]	[3.260]	[3.189]	[2.466]	[3.734]
Observations	2,971	2,925	2,971	2,921	2,971	2,971	2,916	2,971
Pseudo R2	0.757	0.76	0.753	0.756	0.756	0.759	0.755	0.759

Robust z statistics in brackets; * significant at 5%; ** significant at 1%. Logit estimation, based on at most 70 countries from 1945 to 1999, standard errors clustered at the country level. The columns report marginal effects.

All columns include controls for lagged war, log lagged GDP per capita, log lagged population, fraction mountainous terrain, non-contiguous country dummy, oil dummy, new state dummy, instability dummy, democracy lagged (Polity 2), Latin America and Caribbean dummy, Sub-Saharan Africa dummy, East and Southeast Asia dummy.

Question categories are defined as follows: A: Perceptions of Life (42 questions); B: Environment (4 questions); C: Work (25 questions); D: Family (12 questions); E: Politics and Society (59 questions); F: Religion and Morale (30 questions); G: National Identity (3 questions); No G: All questions except category G (172 questions).

Panel B – By Question Type

	Binary	Scale	Multi
	(49q)	(100q)	(26q)
Cultural Fractionalization	-0.334**	-0.257*	-0.058
	[-4.355]	[-2.011]	[-0.578]
Ethnolinguistic Fractionalization	-0.021	-0.021	-0.017
	[-1.318]	[-1.040]	[-0.900]
Chi-square	1.311**	0.562**	0.585*
	[4.213]	[2.949]	[2.573]
Observations	2,971	2,971	2,925
Pseudo R-squared	0.763	0.757	0.762

Robust z statistics in brackets; * significant at 5%; ** significant at 1%. Logit estimation, based on at most 70 countries from 1945 to 1999, standard errors clustered at the country level. The columns report marginal effects.

All columns include controls for lagged war, log lagged GDP per capita, log lagged population, fraction mountainous terrain, non-contiguous territory dummy, oil dummy, new state dummy, instability dummy, democracy lagged (Polity 2), Latin America and Caribbean dummy, Sub-Saharan Africa dummy, East and Southeast Asia dummy.

Appendix Table B37 – Robustness Tests
(Dependent Variable: Incidence of Civil Conflict)

	(1) PRIO25	(2) Expanded Questions	(3) 9 Questions
Cultural Fractionalization	-0.894** [-3.270]	-0.323** [-2.804]	-0.217** [-2.794]
Ethnolinguistic Fractionalization	-0.058 [-1.345]	-0.035 [-1.659]	-0.030 [-1.633]
Chi-square	1.683** [4.410]	0.855** [3.711]	1.293** [3.446]
Observations	2,834	2,921	3,078
Pseudo R-squared	0.591	0.757	0.756

Robust z statistics in brackets; * significant at 5%; ** significant at 1%. Logit estimation, based on at most 72 countries from 1945 to 1999, standard errors clustered at country level. The columns report marginal effects.

All columns include controls for lagged war, log lagged GDP per capita, log lagged population, fraction mountainous terrain, non-contiguous territory dummy, oil dummy, new state dummy, instability dummy, democracy lagged (Polity 2), Latin America and Caribbean dummy, Sub-Saharan Africa dummy, East and Southeast Asia dummy.

Appendix Table B38 - Onset of Civil Conflict and Diversity
(Dependent Variable: Onset of Civil Conflict)

	(1) Onset	(2) Onset	(3) Onset	(4) Onset	(5) Onset	(6) Onset
Cultural Fractionalization	-0.024 [-0.62]			-0.075 [-1.56]	-0.120* [-2.25]	-0.090 [-1.91]
Ethnolinguistic Fractionalization		0.005 [0.94]		-0.009 [-1.11]	-0.008 [-0.92]	-0.008 [-1.11]
Chi-square			0.088 [1.82]	0.175* [2.05]	0.254** [2.95]	0.179* [2.12]
Lagged War	-0.006** [-3.13]	-0.006** [-3.22]	-0.007** [-3.27]	-0.007** [-3.32]	-0.008** [-3.56]	-0.007** [-3.05]
Log Lagged GDP per capita	-0.005* [-2.49]	-0.005** [-3.17]	-0.004** [-2.85]	-0.003 [-1.50]	-0.001 [-0.58]	-0.002 [-1.14]
Log Lagged Population	0.003** [3.55]	0.003** [3.79]	0.003** [4.06]	0.003** [3.69]	0.004** [3.70]	0.003** [3.49]
% Mountainous Terrain	0.000** [2.94]	0.000** [2.74]	0.000** [2.92]	0.000** [2.97]	0.000 [1.73]	0.000** [2.77]
Non Contiguous	0.005 [1.08]	0.004 [0.94]	0.004 [0.92]	0.006 [1.06]	0.006 [1.15]	0.005 [1.01]
Oil	0.021** [2.80]	0.021** [2.73]	0.025** [2.96]	0.024** [2.79]	0.018** [2.65]	0.019* [2.09]
New State	0.075* [1.98]	0.077* [1.99]	0.080* [2.01]	0.074 [1.91]	0.081 [1.88]	0.060 [1.82]
Instability	0.008* [2.00]	0.008* [2.04]	0.008* [2.05]	0.008 [1.93]	0.009* [2.10]	0.006 [1.70]
Democracy Lagged (Polity 2)	0.000 [0.42]	0.000 [0.26]	-0.000 [-0.04]	0.000 [0.02]	0.000 [0.41]	0.000 [0.08]
UK Legal Origin					-0.006 [-0.75]	
French Legal Origin					0.003 [0.26]	
Socialist Legal Origin					0.002 [0.17]	
GDP Growth						-0.027* [-2.30]
GDP Growth Lagged						-0.008 [-0.70]
Observations	2,921	2,921	2,921	2,921	2,705	2,850
Pseudo R-squared	0.147	0.148	0.153	0.158	0.157	0.165

Robust z statistics in brackets; * significant at 5%; ** significant at 1%. Logit estimation, based on at most 69 countries from 1945 to 1999, standard errors clustered at country level. The columns report marginal effects. All columns include dummies for Latin America and Caribbean, Sub-Saharan Africa, East and Southeast Asia.

Appendix Table B39 - Incidence of Civil Conflict and Diversity: Polarization, Genetic Diversity and Distance-based Measures

	(1) Genetic Diversity	(2) F_{ST}	(3) Polarization	(4) Distance	(5) Distance Scale Qs
Cultural	-0.413**	-0.361**	-1.403**	-0.436*	-0.149
Diversity[a]	[-3.415]	[-3.202]	[-3.042]	[-2.487]	[-0.627]
Ethnolinguistic	-0.038	-0.038	-0.619	-0.028	-0.021
Diversity[b]	[-1.779]	[-1.848]	[-1.821]	[-1.406]	[-0.938]
Overlap Measure[c]	0.830**	2.021**	3.630	1.407**	1.221**
	[3.660]	[4.340]	[1.904]	[3.275]	[2.612]
Genetic Diversity	0.432*				
	[2.037]				
Observations	2,870	2,921	2,921	2,921	2,971
R-squared (or Pseudo)	0.759	0.758	0.756	0.757	0.756

Robust z statistics in brackets; * significant at 5%; ** significant at 1%. Logit estimation, based on at most 69 countries from 1945 to 1999, standard errors clustered at country level. The columns report marginal effects.

All columns include controls for lagged war, log lagged GDP per capita, log lagged population, fraction mountainous terrain, non-contiguous territory dummy, oil dummy, new state dummy, instability dummy, democracy lagged (Polity 2), Latin America and Caribbean dummy, Sub-Saharan Africa dummy, East and Southeast Asia dummy.

a: Cultural fractionalization in columns (1) and (2), cultural polarization in column (3) and cultural fractionalization with distances in columns (4) and (5).

b: Ethnolinguistic fractionalization in columns (1), (2), (4) and (5), ethnolinguistic polarization in column (3).

c: Chi-square in column (1), F_{ST} in column (2), overlap-polarization in column (3), F_{ST} with distances in columns (4) and (5).

Column (4) is based on all questions, whereas column (5) is based on only scale questions.

Appendix Table B40 - Incidence of Civil Conflict and Diversity: Pseudo R-Squared, with Chi-square

	(1) Incidence	(2) Incidence	(3) Incidence	(4) Incidence	(5) Incidence	(6) Incidence	(7) Incidence	(8) Incidence
Cultural Fractionalization	-0.391** [-3.330]		-1.930** [-2.648]		-0.926 [-1.412]			-1.651** [-2.931]
Ethnolinguistic Fractionalization	-0.037 [-1.769]		-0.145 [-1.358]			0.094 [1.209]		-0.006 [-0.060]
Chi-square	0.870** [3.823]		3.258** [2.943]				1.476* [1.974]	2.616* [2.116]
Lagged War	0.840** [27.877]	0.864** [41.867]						
Observations	2,921	4,772	2,921	4,772	2,921	2,921	2,921	3,007
Pseudo R-squared (McFadden)	0.758	0.749	0.258	0.199	0.202	0.194	0.209	0.112
Pseudo R-squared (Cox & Snell)	0.461	0.434	0.190	0.140	0.152	0.147	0.157	0.086

Robust z statistics in brackets; * significant at 5%; ** significant at 1%

Logit estimation, based on at most 111 countries from 1945 to 1999, standard errors clustered at country level. The columns report marginal effects. All columns except column (8) include controls for log lagged GDP per capita, log lagged population, fraction mountainous terrain, non-contiguous territory dummy, oil dummy, new state dummy, instability dummy, democracy lagged (Polity 2), Latin America and Caribbean dummy, Sub-Saharan Africa dummy, East and Southeast Asia dummy. Column (8) contains no additional controls.

Appendix Table B41 – Incidence of Civil Conflict and Diversity: Pseudo R-Squared, with F_{ST}

	(1) Incidence	(2) Incidence	(3) Incidence	(4) Incidence	(5) Incidence
Cultural Fractionalization	-0.361** [-3.202]		-0.187 [-1.601]		
Ethnolinguistic Fractionalization	-0.038 [-1.848]				0.019 [1.092]
F_{ST}	2.021** [4.340]			1.153** [3.020]	
Lagged War	0.842** [26.740]	0.864** [41.466]	0.860** [30.660]	0.863** [32.387]	0.868** [33.743]
Log Lagged GDP per capita	0.002 [0.536]	-0.009* [-2.263]	-0.005 [-0.895]	-0.006 [-1.443]	-0.008 [-1.736]
Log Lagged Population	0.012** [4.167]	0.010** [4.688]	0.011** [3.687]	0.013** [4.342]	0.012** [3.979]
Fraction Mountainous	0.046** [2.836]	0.023* [2.022]	0.041* [2.532]	0.043* [2.442]	0.040* [2.341]
Non Contiguous	0.028 [1.633]	0.025 [1.879]	0.025 [1.607]	0.016 [1.204]	0.017 [1.273]
Oil	0.019 [1.468]	0.014 [1.365]	0.020 [1.561]	0.023 [1.705]	0.021 [1.497]
New State	0.213* [2.151]	0.143* [2.410]	0.200* [2.144]	0.241* [2.351]	0.222* [2.287]
Instability	-0.009 [-0.987]	-0.008 [-1.201]	-0.009 [-1.001]	-0.009 [-0.957]	-0.009 [-0.941]
Democracy Lagged (Polity 2, range -1 to 1)	0.003 [0.460]	0.000 [0.040]	0.006 [0.902]	0.001 [0.187]	0.004 [0.561]
Latin America and Caribbean	0.056 [1.886]	0.023 [1.789]	0.017 [0.983]	0.019 [1.009]	0.007 [0.482]
Sub-Saharan Africa	0.057 [1.624]	0.019 [1.699]	0.022 [0.898]	0.003 [0.214]	0.001 [0.052]
East and Southeast Asia	-0.014 [-1.673]	-0.009 [-1.166]	-0.008 [-0.680]	-0.017* [-2.203]	-0.010 [-1.124]
Observations	2,921	4,772	2,921	2,921	2,921
Pseudo R-squared (McFadden)	0.758	0.749	0.752	0.754	0.752

Robust z statistics in brackets; * significant at 5%; ** significant at 1%

Logit regressions, based on maximum 111 countries from 1945 to 1999

Appendix Table B42 - Incidence of Civil Conflict and Diversity: Robustness to Ethnic and Linguistic Classification (Dependent Variable: Incidence of Civil Conflict)

	(1)	(2) Without Latin America	(3) Without Sub-Saharan Africa
Cultural Fractionalization	**-0.383** [-3.195]	**-0.347** [-2.825]	**-0.414** [-3.888]
Ethnolinguistic Fractionalization	**-0.046** [-1.368]	**-0.037** [-1.529]	**-0.045*** [-2.289]
Chi-square	**0.811** [2.825]	**0.922** [3.584]	**0.833** [3.802]
CF * Ethnic Classification	-0.022 [-0.656]		
ELF * Ethnic Classification	0.015 [0.385]		
Chi Square * Ethnic Classification	0.230 [0.576]		
Lagged War	0.836** [27.060]	0.822** [25.455]	0.819** [20.562]
Log Lagged GDP per capita	0.003 [0.512]	0.003 [0.684]	0.003 [0.712]
Log Lagged Population	0.013** [3.956]	0.014** [4.489]	0.011** [3.833]
Fraction Mountainous	0.033* [1.973]	0.035 [1.931]	0.037* [2.103]
Non Contiguous	0.023 [1.514]	0.020 [1.396]	0.026 [1.624]
Oil	0.018 [1.532]	0.033 [1.869]	0.024 [1.625]
New State	0.210* [2.172]	0.232* [2.162]	0.260* [2.061]
Instability	-0.008 [-1.004]	-0.008 [-0.765]	-0.007 [-0.836]
Democracy Lagged (Polity 2)	0.004 [0.586]	0.004 [0.460]	0.003 [0.473]
Latin America and Caribbean	0.064 [1.839]		0.065* [2.099]
Sub-Saharan Africa	0.049 [1.315]	0.048 [1.505]	
East and Southeast Asia	-0.019** [-2.671]	-0.018* [-2.028]	-0.014 [-1.686]
Observations	2,921	2,449	2,542
R-squared (or Pseudo)	0.759	0.757	0.771

Robust z statistics in brackets; * significant at 5%; ** significant at 1%. Logit estimation, based on at most 69 countries from 1945 to 1999, standard errors clustered at the country level. Columns report marginal effects. Column (1) includes interaction terms between the different diversity measures and a dummy that takes a value of 1 when the group classification is based on ethnicity (variable x051) and a value of 0 when the group classification is based on the respondent's language (variable g016). Column (2) excludes Latin America and Column (3) excludes sub-Saharan Africa.

Appendix Table B43 – Incidence of Civil Conflict and Diversity: Alternative Estimation Methods

	(1) Incidence Logit	(2) Share All Period Linear	(3) Incidence Rare Events	(4) Incidence Linear	(5) Incidence Decade Logit
Cultural	-1.930**	-2.061*	-14.406**	-0.362**	-1.679
Fractionalization	[-2.648]	[-2.022]	[-2.888]	[-3.300]	[-1.767]
Ethnolinguistic	-0.145	-0.113	-1.351	-0.027	-0.041
Fractionalization	[-1.358]	[-0.796]	[-1.696]	[-1.653]	[-0.247]
Chi-square	3.258**	3.782*	32.130**	0.760**	3.644*
	[2.943]	[2.518]	[3.392]	[3.641]	[2.239]
Lagged War			5.975**	0.873**	0.497**
			[22.749]	[48.754]	[4.597]
Log Lagged GDP	0.012	-0.112*	0.104	-0.003	-0.005
per capita[a]	[0.625]	[-2.464]	[0.433]	[-0.810]	[-0.096]
Log Lagged Population[a]	0.042**	0.042*	0.426**	0.009**	0.028
	[3.141]	[2.202]	[3.796]	[3.549]	[1.306]
Fraction Mountainous	0.154*	0.167	1.497*	0.036*	0.331**
	[2.079]	[1.632]	[2.107]	[2.458]	[2.732]
Non Contiguous	0.117	0.105	0.765*	0.018	0.136
	[1.182]	[1.376]	[2.214]	[1.685]	[1.571]
Oil	0.069	0.114	0.638	0.014	0.075
	[1.227]	[1.557]	[1.370]	[1.602]	[0.995]
New State	0.063	-0.188	2.431**	0.101*	-0.127*
	[0.764]	[-0.915]	[4.294]	[2.603]	[-2.259]
Instability	0.025	0.020	-0.366	-0.014	0.129
	[0.866]	[0.110]	[-0.956]	[-1.230]	[1.937]
Democracy Lagged	0.010	0.106*	0.119	0.005	0.008
(Polity 2)[a]	[0.576]	[2.429]	[0.485]	[1.041]	[0.181]
Latin America	0.239	0.090	1.323**	0.020*	0.050
and Caribbean	[1.526]	[1.289]	[2.685]	[2.265]	[0.442]
Sub-Saharan Africa	0.184	-0.066	1.184*	0.010	0.146
	[1.268]	[-0.611]	[2.126]	[0.769]	[0.799]
East and Southeast	-0.044	-0.106	-0.695	-0.016	-0.137*
Asia	[-1.091]	[-1.358]	[-1.615]	[-1.538]	[-2.127]
Constant		1.570*		0.116	
		[2.498]		[1.782]	
Observations	2,921	69	2,921	2,921	216
Number of clusters	69	69	69	69	56
R-squared (or Pseudo)	0.258	0.519		0.8069	0.405

Robust z statistics in brackets; * significant at 5%; ** significant at 1%

Based on countries from 1945 to 1999

Column 1: logit estimation without lagged dependent variable; Column 2: linear estimation, one observation per country, dependent variable is share of years at war; Column 3: rare events logit estimation; Column 4: linear estimation; Column 5: logit estimation based on decades.

a: lagged variables except in Column 2 where they refer to means over the entire period

Appendix Table B44 – Incidence of Civil Conflict and Diversity: Ethnolinguistic vs Regional

	(1) Incidence Ethnic	(2) Incidence Regional	(3) Incidence Ethnic & Regional
Cultural Fractionalization	**-0.391**** [-3.330]	**-0.395**** [-3.871]	**-0.617**** [-4.900]
Ethnolinguistic Fractionalization	**-0.037** [-1.769]		**-0.067**** [-2.795]
Ethnolinguistic Chi-square	**0.870**** [3.823]		**1.257**** [4.331]
Regional Fractionalization		-0.011 [-0.348]	0.059 [1.051]
Regional Chi-square		0.103 [1.358]	-0.162 [-1.603]
Observations	2,921	3,417	2,809
R-squared (or Pseudo)	0.758	0.767	0.758

Robust z statistics in brackets; * significant at 5%; ** significant at 1%

Based on countries from 1945 to 1999

All columns include controls for lagged war, log lagged GDP per capita, log lagged population, fraction mountainous terrain, non-contiguous territory dummy, oil dummy, new state dummy, instability dummy, democracy lagged (Polity 2), Latin America and Caribbean dummy, Sub-Saharan Africa dummy, East and Southeast Asia dummy.

Appendix Table B45 – Incidence of Civil Conflict and Diversity: Quadratic Terms

	(1) CF & ELF Squared	(2) All Squared
Cultural Fractionalization	2.389 [1.038]	2.354 [1.022]
Cultural Fractionalization (Squared)	-2.767 [-1.200]	-2.724 [-1.180]
Ethnolinguistic Fractionalization	-0.001 [-0.020]	-0.004 [-0.062]
Ethnolinguistic (Squared) Fractionalization	-0.043 [-0.530]	-0.041 [-0.496]
Chi Square	0.925** [4.102]	1.123* [2.211]
Chi Square (Squared)		-1.78 [-0.423]
Observations	2921	2921
R-squared (or Pseudo)	0.759	0.759

Robust z statistics in brackets; * significant at 5%; ** significant at 1%
Logit estimation, based on at most 69 countries from 1945 to 1999, standard errors clustered at country level.
All columns include controls for lagged war, log lagged GDP per capita, log lagged population, fraction mountainous terrain, non-contiguous territory dummy, oil dummy, new state dummy, instability dummy, democracy lagged (Polity 2), Latin America and Caribbean dummy, Sub-Saharan Africa dummy, East and Southeast Asia dummy.

Table B46: Public Goods Provision and Diversity

Robustness to the use of more public goods indicators and to the questions set used to compute diversity

	(1) PCA of 10 public goods	(2) Expanded questions	(3) Cat A	(4) Cat B	(5) Cat C	(6) Cat D	(7) Cat E	(8) Cat F	(9) Cat G
Cultural Fractionalization	19.581* [2.146]	18.005* [2.484]	-0.682 [-0.106]	3.257 [1.276]	5.929* [2.013]	19.732** [3.398]	16.946 [1.843]	7.538** [4.156]	2.182 [1.298]
Chi-square	-34.005 [-1.997]	-27.816* [-2.461]	-14.625 [-0.925]	-22.210** [-3.100]	-26.230** [-2.726]	-26.212* [-2.377]	-27.526* [-2.301]	-16.765* [-2.552]	4.880 [1.021]
Ethnolinguistic Fractionalization	1.817 [1.437]	1.133 [1.244]	0.323 [0.285]	1.349 [1.964]	1.315 [1.615]	1.391 [1.278]	1.051 [1.040]	0.579 [0.773]	-0.844 [-0.938]
Absolute Latitude	0.095** [3.830]	0.058** [3.103]	0.082** [4.764]	0.079** [4.149]	0.069** [3.787]	0.057** [3.425]	0.072** [4.167]	0.037* [2.442]	0.079** [4.179]
Constant	-11.579* [-2.626]	-9.804* [-2.618]	-1.062 [-0.327]	-2.937 [-1.621]	-3.946 [-1.912]	-10.739** [-3.494]	-10.928 [-1.895]	-3.109** [-2.787]	-2.766* [-2.081]
Observations	42	59	60	60	60	59	60	60	59
R-squared	0.748	0.752	0.687	0.756	0.7459	0.783	0.733	0.781	0.698

Robust t-statistics in brackets; * significant at 5%; ** significant at 1%.

This table shows results from OLS regressions of the first principal component of 10 public goods measures (column 1) or 8 public goods measures (columns 2-9) on diversity measures and a set of controls.

All regressions include the following controls in addition to those reported: East and Southeast Asia dummy, Sub-Saharan Africa dummy, Latin America and Caribbean dummy, French legal origin, German legal origin, Scandinavian legal origin.

Column (1) considers the determinants of the first principal component of all 10 public goods measures (Sample size falls to 42).

Column (2) considers the effect of diversity measures when CF and Chi-square are computed over an expanded set of World Values Survey questions (where we require that the questions be asked in at least 30 countries rather than 50 as we do in the rest of the paper).

Columns (3) - (9) compute diversity over different WVS question categories (A through G) to see if any given category drives the results when averaging over all questions.

Appendix Table B47: Public Goods Provision and Diversity

Robustness to questions set used to compute diversity, to the functional form of diversity measures, and to the criterion used to define identity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Binary	Scale	Multi	Persistent	Non-persistent	Polarization	Distance	Ethnic	Language
Cultural Diversity	15.565**	22.816**	19.629**	26.694*	13.341**	42.924	25.680*	1.635	-0.213
	[2.840]	[2.982]	[2.988]	[2.556]	[2.974]	[1.129]	[2.397]	[0.855]	[-0.206]
Overlap Measure	-53.674	-26.219*	-21.083	-32.449**	-28.911*	5.544	-55.484	22.648*	21.509*
	[-1.991]	[-2.675]	[-1.767]	[-2.770]	[-2.356]	[0.073]	[-1.937]	[2.313]	[2.451]
Ethnolinguistic Diversity	1.266	0.777	1.014	1.499	0.939	1.231	1.278	-18.200	-37.956*
	[1.196]	[0.914]	[1.253]	[1.690]	[1.061]	[0.092]	[1.341]	[-1.114]	[-2.487]
Absolute latitude	0.054**	0.050**	0.052*	0.056**	0.043*	0.072*	0.064**	0.040	0.075*
	[3.001]	[2.861]	[2.573]	[3.109]	[2.258]	[2.657]	[3.327]	[1.135]	[2.553]
Constant	-5.674**	-13.149**	-11.700**	-16.941*	-4.806**	-9.452	-9.167*	-11.748*	-11.460*
	[-2.798]	[-3.024]	[-3.115]	[-2.613]	[-2.796]	[-1.619]	[-2.602]	[-2.095]	[-2.734]
Observations	60	60	60	60	59	59	59	32	27
R-squared	0.746	0.772	0.766	0.757	0.761	0.711	0.7497	0.700	0.9007

Robust t-statistics in brackets; * significant at 5%; ** significant at 1%.

This table shows results from OLS regressions of the first principal component of 8 public goods measures on diversity measures and a set of controls. All regressions include the following controls in addition to those reported: East and Southeast Asia dummy, Sub-Saharan Africa dummy, Latin America and Caribbean dummy, French legal origin, German legal origin, Scandinavian legal origin.

Columns (1) – (3) break down the questions set used to compute diversity measures by question type (binary, scale, multiple unordered answers). Columns (4) and (5) break down the set of WVS questions by whether they display persistent patterns of cultural fractionalization over time or not. Columns (6) and (7) measure diversity using different functional forms of the diversity measures, based on, respectively, polarization and the consideration of distances between scale questions (see Section 3.4 of the main text). Other columns include measures based on fractionalization. Columns (7) and (8) breaks down the sample according to whether the main criterion used to define identity is language or ethnicity (see Section 2.2 of the main text).

68

Appendix Table B48: Public Goods Provision and Diversity
Breakdown by type of public good, parsimonious specification

	Dependent Variable:	Cultural Fractionalization	Chi-square	Ethnolinguistic Fractionalization	Observations	R-squared
(1)	Child survival	0.275**	-0.330**	0.005	70	0.798
		[3.064]	[-2.736]	[0.488]		
(2)	Immunization	37.831	-102.642	1.852	70	0.492
		[0.717]	[-1.164]	[0.256]		
(3)	Hospital beds	22.328*	-32.148	1.835	71	0.635
		[2.502]	[-1.823]	[1.043]		
(4)	Literacy	177.782*	-301.970**	13.688	57	0.668
		[2.366]	[-3.026]	[1.220]		
(5)	School attainment	3.204**	-5.176**	0.221	64	0.624
		[2.913]	[-3.502]	[1.669]		
(6)	Sanitation	251.663**	-366.659*	17.812	69	0.723
		[3.048]	[-2.045]	[1.481]		
(7)	Water	79.229*	-135.655*	9.808	70	0.795
		[2.555]	[-2.346]	[1.771]		
(8)	Infrastructure	11.288*	-16.110**	1.034	70	0.555
		[2.632]	[-3.164]	[1.637]		
(9)	Road density	48.911	-44.402	4.298	71	0.438
		[1.047]	[-0.533]	[0.759]		
(10)	Rail density	0.576	0.159	-0.046	66	0.732
		[0.761]	[0.111]	[-0.262]		

Robust t-statistics in brackets; * significant at 5%; ** significant at 1%.

This table shows results from OLS regressions of each of 10 public goods measures on diversity measures and a set of controls.

All regressions include the following controls in addition to those reported: East and Southeast Asia dummy, Sub-Saharan Africa dummy, Latin America and Caribbean dummy, French legal origin, German legal origin, Scandinavian legal origin, absolute latitude and a constant.

Appendix Table B49: Public Goods Provision and Diversity

Breakdown by type of public good, expanded specification including a set of geographic controls

	Dependent Variable:	Cultural Fractionalization	Chi-square	Ethnolinguistic Fractionalization	Observations	R-squared
(1)	Child survival	0.343** [3.877]	-0.336** [-2.839]	0.006 [0.610]	68	0.841
(2)	Immunization	96.337 [1.689]	-152.098 [-1.828]	1.852 [0.258]	68	0.608
(3)	Hospital beds	17.380 [1.618]	-40.181* [-2.579]	2.183 [1.245]	68	0.705
(4)	Literacy	172.683 [2.020]	-297.443** [-3.312]	16.738 [1.433]	54	0.740
(5)	School attainment	3.749** [3.070]	-5.793** [-3.087]	0.288 [1.787]	61	0.668
(6)	Sanitation	330.321** [4.425]	-395.242* [-2.638]	14.623 [1.435]	66	0.808
(7)	Water	110.200** [3.838]	-128.968* [-2.556]	10.379* [2.018]	67	0.851
(8)	Infrastructure	13.545** [2.717]	-13.365 [-1.947]	0.600 [0.756]	66	0.605
(9)	Road density	52.749 [1.205]	-20.039 [-0.234]	3.274 [0.534]	68	0.480
(10)	Rail density	-0.249 [-0.268]	1.167 [0.858]	-0.107 [-0.698]	66	0.770

Robust t-statistics in brackets; * significant at 5%; ** significant at 1%.

This table shows results from OLS regressions of each of 10 public goods measures on diversity measures and a set of controls.

All regressions include the following controls in addition to those reported: East and Southeast Asia dummy, Sub-Saharan Africa dummy, Latin America and Caribbean dummy, French legal origin, German legal origin, Scandinavian legal origin, absolute latitude, soil fertility, terrain roughness, mean elevation, dummy for island, log land area, dummy for landlocked country, and a constant.

Appendix Table B50: Income and Diversity Measures,
(Dependent variable: log per capita income 2005)

	(1)	(2)	(3)	(4)	(5)	(6)
	No controls	Add geographic controls	Add region dummies	Add legal origins	Add religion shares	Add quadratic terms
Cultural ELF	11.529	10.494	12.732	11.480	10.592	46.287
	(4.147)**	(6.420)**	(6.381)**	(4.949)**	(4.449)**	(1.180)
Cultural ELF, squared						-36.094
						(0.916)
Chi-square	-9.999	-10.444	-12.898	-15.874	-13.799	-26.830
	(2.325)*	(2.425)*	(3.079)**	(4.215)**	(3.592)**	(2.543)*
Chi-square, squared						123.096
						(1.229)
Ethnic ELF	-0.725	0.304	0.666	0.790	0.630	2.180
	(1.578)	(0.726)	(1.673)	(2.467)*	(2.013)*	(1.844)
ELF, squared						-1.785
						(1.216)
Absolute latitude		0.032	0.017	0.007	0.002	0.003
		(4.355)**	(1.939)	(0.940)	(0.295)	(0.323)
% land area in the tropics		0.012	-0.083	-0.473	-0.628	-0.880
		(0.036)	(0.353)	(1.662)	(1.734)	(2.606)*
Landlocked dummy		-0.714	-0.454	-0.225	-0.171	-0.095
		(3.156)**	(2.043)*	(1.118)	(1.007)	(0.513)
Island dummy		0.992	0.732	0.607	0.662	0.841
		(4.730)**	(4.014)**	(3.048)**	(2.793)**	(3.455)**
Constant	3.618	2.678	2.114	3.345	3.712	-5.154
	(2.409)*	(3.063)**	(2.509)*	(3.092)**	(3.127)**	(0.528)
R^2	0.28	0.68	0.78	0.83	0.87	0.88
N	73	72	72	71	69	69

Robust t-statistics in parentheses; * $p<0.05$; ** $p<0.01$

This table displays results from an OLS regression of log per capita income in 2005 on diversity measures and a set of controls.

Column (3) add regional dummies as follows: Southeast Asia Dummy, Sub-Saharan Africa Dummy, Latin America and Caribbean Dummy (estimates not reported).

Column (4) includes the same regressors as column (3) plus: UK legal origin, French legal origin, Scandinavian legal origin (estimates not reported).

Column (5) includes the same regressors as column (4) plus: Share of Protestants, share of Roman Catholics, share of Muslims (estimates not reported).

Column (6) includes the same regressors as column (5) plus quadratic terms in ELF, Chi-square and CF to test for any nonlinear effects.

Appendix B6: Diversity Measures and Questions Included in the Study

Appendix Table B51 – Indices of Cultural Diversity, Ethnic Diversity and Overlap

Iso Code	Country	Cultural fractionalization (CF)	Ethnolinguistic fractionalization (ELF)	FST	Chi-Square
ALB	Albania	0.5239	0.0269	0.0017	0.0055
DZA	Algeria	0.4856	0.2639	0.0055	0.0175
AND	Andorra	0.5409	0.6166	0.0103	0.0267
ARG	Argentina	0.5381	0.1313	0.0032	0.0090
ARM	Armenia	0.5244	0.0643	0.0038	0.0112
AUS	Australia	0.5524	0.2441	0.0058	0.0150
AZE	Azerbaijan	0.4930	0.2495	0.0152	0.0363
BGD	Bangladesh	0.4356	0.1507	0.0038	0.0100
BLR	Belarus	0.5257	0.3833	0.0064	0.0155
BIH	Bosnia Herzogovina	0.5542	0.6261	0.0151	0.0307
BRA	Brazil	0.5632	0.5820	0.0052	0.0132
BGR	Bulgaria	0.5377	0.2955	0.0104	0.0230
BFA	Burkina Faso	0.5427	0.6674	0.0208	0.0588
CAN	Canada	0.5583	0.6256	0.0119	0.0254
CHL	Chile	0.5650	0.1621	0.0042	0.0118
CHN	China	0.4939	0.1356	0.0126	0.0357
CYP	Cyprus	0.5830	0.5038	0.0306	0.0688
CZE	Czech Republic	0.5374	0.0071	0.0025	0.0071
EGY	Egypt	0.4299	0.6350	0.0070	0.0156
EST	Estonia	0.5174	0.4862	0.0228	0.0462
ETH	Ethiopia	0.5453	0.6940	0.0195	0.0576
FIN	Finland	0.5519	0.0531	0.0040	0.0122
FRA	France	0.5825	0.1352	0.0086	0.0294
GEO	Georgia	0.4947	0.2502	0.0078	0.0191
GER	Germany	0.5760	0.0729	0.0028	0.0093
GHA	Ghana	0.5293	0.5820	0.0162	0.0446
GBR	Great Britain	0.5728	0.1274	0.0090	0.0278
GTM	Guatemala	0.5129	0.4093	0.0023	0.0064
IND	India	0.5655	0.8517	0.0588	0.1281
IDN	Indonesia	0.4616	0.5956	0.0167	0.0379
IRN	Iran	0.4998	0.6136	0.0080	0.0211
IRQ	Iraq	0.4892	0.2960	0.0227	0.0415
ITA	Italy	0.5574	0.0000	0.0000	0.0000
JPN	Japan	0.5273	0.0043	0.0009	0.0028
JOR	Jordan	0.4273	0.4976	0.0066	0.0165
KGZ	Kyrgyzstan	0.5489	0.5860	0.0193	0.0420
LVA	Latvia	0.5302	0.5869	0.0173	0.0469
LTU	Lithuania	0.5031	0.1697	0.0042	0.0098
MKD	Macedonia	0.5428	0.4398	0.0344	0.0674
MYS	Malaysia	0.5632	0.6528	0.0336	0.0919
MLI	Mali	0.5657	0.4331	0.0128	0.0360

Iso Code	Country	Cultural fractionalization (CF)	Ethnolinguistic fractionalization (ELF)	FST	Chi-Square
MEX	Mexico	0.5476	0.6426	0.0057	0.0148
MDA	Moldova	0.5450	0.3695	0.0108	0.0282
MAR	Morocco	0.4452	0.2235	0.0031	0.0084
NZL	New Zealand	0.5379	0.0408	0.0065	0.0227
NGA	Nigeria	0.4987	0.7675	0.0103	0.0231
NOR	Norway	0.5394	0.0661	0.0028	0.0092
PAK	Pakistan	0.4451	0.7562	0.0198	0.0482
PER	Peru	0.5234	0.5767	0.0081	0.0219
PHL	Philippines	0.5168	0.7655	0.0295	0.0765
POL	Poland	0.5443	0.0199	0.0021	0.0060
PRI	Puerto Rico	0.5059	0.6382	0.0054	0.0169
ROM	Romania	0.5229	0.1387	0.0049	0.0145
RUS	Russian Federation	0.5388	0.4837	0.0035	0.0098
SAU	Saudi Arabia	0.5082	0.5148	0.0137	0.0345
SER	Serbia	0.5869	0.1896	0.0102	0.0299
SGP	Singapore	0.5099	0.6778	0.0321	0.0629
SVK	Slovak Republic	0.5418	0.1339	0.0100	0.0305
SVN	Slovenia	0.5469	0.0559	0.0020	0.0062
ZAF	South Africa	0.5552	0.8389	0.0191	0.0445
KOR	South Korea	0.5518	0.0000	0.0000	0.0000
ESP	Spain	0.5522	0.2944	0.0052	0.0125
SWE	Sweden	0.5438	0.2277	0.0086	0.0277
CHE	Switzerland	0.5566	0.6380	0.0139	0.0312
TWN	Taiwan	0.516	0.5394	0.0089	0.0203
TZA	Tanzania	0.4651	0.2744	0.0044	0.0149
THA	Thailand	0.5611	0.6979	0.0495	0.1218
TTO	Trinidad and Tobago	0.5325	0.6209	0.0094	0.0270
TUR	Turkey	0.5186	0.2041	0.0092	0.0253
UGA	Uganda	0.5166	0.8114	0.0185	0.0469
UKR	Ukraine	0.5429	0.5018	0.0062	0.0135
URY	Uruguay	0.5437	0.1126	0.0041	0.0111
USA	USA	0.5512	0.3359	0.0067	0.0169
VEN	Venezuela	0.5454	0.6602	0.0050	0.0156
VNM	Vietnam	0.5039	0.0814	0.0016	0.0043
ZMB	Zambia	0.6024	0.8045	0.0293	0.0834

Appendix Table B52 – List of Questions Asked in at Least 50 Countries, by Category and Type

A: Perceptions of Life	
Binary questions	
a025	love and respect parents always/earned
a029	learn children at home: independence
a030	learn children at home: hard work
a032	learn children at home: feeling of responsibility
a034	learn children at home: imagination
a035	learn children at home: tolerance+respect
a038	learn children at home: thrift
a039	learn children at home: determination/perseverance
a040	learn children at home: religious faith
a041	learn children at home: unselfishness
a042	learn children at home: obedience
a124_01	don't like as neighbours: people with criminal record
a124_02	don't like as neighbours: people of different race
a124_03	don't like as neighbours: heavy drinkers
a124_04	don't like as neighbours: emotionally unstable people
a124_06	don't like as neighbours: immigrants/foreign workers
a124_07	don't like as neighbours: people with AIDS
a124_08	don't like as neighbours: drug addicts
a124_09	don't like as neighbours: homosexuals
a124_12	neighbours: people of a different religion
a165	people can be trusted/can't be too careful
Scale questions	
a001	how important in your life: family
a002	how important in your life: friends and acquaintances
a003	how important in your life: leisure time
a004	how important in your life: politics
a005	how important in your life: work
a006	how important in your life: religion
a008	taking all things together how happy are you
a009	describe your state of health these days
a062	how often discuss politics with friends
a098	active/inactive membership of church or religious organization
a099	active/inactive membership of sport or recreation
a100	active/inactive membership of art, music or educational organization
a101	active/inactive membership of labour unions
a102	active/inactive membership of political party
a103	active/inactive membership of environmental organization
a104	active/inactive membership of professional organization
a105	active/inactive membership of charitable/humanitarian organization
a106	active/inactive membership of any other organization
a170	how satisfied are you with your life
a173	how much control over your life
Unordered response questions	
a026	parents responsibilities to their children at expense of/not sacrifice own well-being

B: Environment	
Scale questions	
b001	environment: giving part of income
b002	environment: increase taxes to prevent environmental pollution
b003	environment: government should reduce environmental pollution
Unordered response questions	
b008	protecting environment vs. economic growth

C: Work	
Binary questions	
c011	important in a job: good pay
c012	important in a job: not too much pressure
c013	important in a job: job security
c014	important in a job: respected job
c015	important in a job: good hours
c016	important in a job: use initiative
c017	important in a job: generous holidays
c018	important in a job: achieving something
c019	important in a job: responsible job
c020	important in a job: interesting job
c021	important in a job: meeting abilities
c059	fair/unfair: quicker secretary is paid more
Scale questions	
c006	how satisfied with financial situation
c008	work compared with leisure
c036	job needed to develop talents
c037	humiliating receiving money without working
c038	people turn lazy not working
c039	work is a duty towards society
c041	work always comes first
Unordered response questions	
c001	jobs are scarce: giving men priority
c002	jobs are scarce: giving...(nation) priority
c009	first choice if looking for a job
c010	second choice if looking for a job
c060	how should business and industry be managed
c061	follow instructions at work/must be convinced first

D: Family	
Binary questions	
d018	children need both parents to grow up happily
d019	women need children in order to be fulfilled
d022	marriage is outdated
Scale questions	
d017	ideal size of a family
d054	one of main goals in life has been to make my parents proud
d055	make effort to live up to what my friends expect

d056	working mother warm relationship with children
d057	being housewife as fulfilling as paid job
d058	husband+wife contribute to household income
d059	men make better political leaders than women do
d060	university is more important for a boy than for a girl
Unordered response questions	
d023	woman wants to have child as single parent without stable relationship with a man

E: Politics and Society	
Binary questions	
e128	country is run by big interest vs. for all people's benefit
Scale questions	
e014	good/bad: less emphasis on money
e015	good/bad: decrease work importance
e016	good/bad: more emphasis on technology
e018	good/bad: more respect for authority
e019	good/bad: more emphasis on family life
e023	how interestedin politics
e025	political action: signing a petition
e026	political action: joining in boycotts
e027	political action: attending lawful demonstrations
e028	political action: joining unofficial strikes
e029	political action: occupying buildings/factories
e033	political view: left-right
e034	opinion on society
e035	equalize incomes vs. incentives for individual effort
e036	private vs. government ownership business
e037	individual vs. state responsibility for providing
e039	competition good vs. harmful for people
e040	hard work brings better life vs. doesn't bring success
e041	accumulate wealth at expense of others vs. enough for everyone
e069_01	how much confidence in: church
e069_02	how much confidence in: armed forces
e069_04	how much confidence in: the press
e069_05	how much confidence in: trade unions
e069_06	how much confidence in: the police
e069_07	how much confidence in: parliament
e069_08	how much confidence in: civil service
e069_10	confidence: television
e069_11	how much confidence in: government
e069_12	how much confidence in: political parties
e069_13	how much confidence in: major companies
e069_14	how much confidence in: environmental organizations
e069_15	confidence: the women´s movement
e069_17	how much confidence in: justice system
e069_20	how much confidence in: United Nations Organisation
e114	political system: strong leader

e115	political system: experts making decisions
e116	political system: the army ruling
e117	political system: democratic
e120	democracy: causes bad economy
e121	democracy: is indecisive
e122	democracy: cannot maintain order
e123	democracy: best political system
e124	how much respect for human rights nowadays
e125	satisfaction with the people in national office
Unordered response questions	
e001	aims of this country: most important
e002	aims of this country: 2nd most important
e003	aims of respondent: first choice
e004	aims of respondent: second choice
e005	aims of respondent II: first choice
e006	aims of respondent II: second choice
e012	are you willing to fight for country
e022	scientific advances help/harm mankind
e135	who should decide: international peacekeeping
e136	who should decide: protection of the environment
e137	who should decide: aid to developing countries
e138	who should decide: refugees
e139	who should decide: human rights
e143	work: people from less developed countries

F: Religion and Morale	
Binary questions	
f024	do you belong to a religious denomination
f035	church answers to: moral problems
f036	church answers to: family life problems
f037	church answers to: spiritual needs
f038	church answers to: social problems
f050	do you believe in: god
f051	do you believe in: life after death
f052	do you believe in: a soul
f053	do you believe in: hell
f054	do you believe in: heaven
f064	do you get comfort and strength from religion
f065	do you take moments of prayer/meditation
Scale questions	
f001	how often: think about meaning of life
f028	how often attend religious services
f063	how important is god in your life
f102	politicians and god
f103	religious leaders and influence voting
f104	religion and public office
f105	religious leaders and influence government decisions

f114	do you justify: claiming state benefits
f115	do you justify: avoiding fare on public transport
f116	do you justify: cheating on tax
f117	do you justify: accepting a bribe
f118	do you justify: homosexuality
f119	do you justify: prostitution
f120	do you justify: abortion
f121	do you justify: divorce
f122	do you justify: euthanasia
f123	do you justify: suicide
Unordered response questions	
f034	are you a religious person

G: Nationality and Identity	
Scale questions	
g006	how proud are you to be a … (country) citizen
Unordered response questions	
g001	belong to geographic group - most
g002	belong to geographic group: next