

Paolo Masella

National identity and ethnic diversity

**Article (Accepted version)
(Refereed)**

Original citation:

Masella, Paolo (2013) *National identity and ethnic diversity*. [Journal of Population Economics](#), 26 (2). pp. 437-454. ISSN 0933-1433

DOI: [10.1007/s00148-011-0398-0](https://doi.org/10.1007/s00148-011-0398-0)

© 2011 [Springer-Verlag](#)

Financial support from the Deutsche Forschungsgemeinschaft through the Collaborative Research Center 884.

This version available at: <http://eprints.lse.ac.uk/59313/>

Available in LSE Research Online: September 2014

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

National Identity and Ethnic Diversity

Paolo Masella*

This Version: October 12, 2011

Abstract

In countries with high levels of ethnic diversity "nation building" has been proposed as a mechanism for integration and conflict reduction. This paper examines the relationship between ethnic diversity and national sentiment. We use individual data from the World Values Survey and, contrary to conventional wisdom, we find no evidence of lower intensity of national sentiment in more ethnically fragmented countries or in minority groups. National feelings in a minority can be higher or lower than in a majority, depending on the degree of ethnic diversity of a country. On the one hand, in countries with high ethnic diversity, nationalist feelings are less strong in minority groups than in the majorities; on the other hand, in countries with low ethnic diversity, the reverse is true.

Keywords: Identity, Ethnic Diversity, Nation-Building

JEL: A14, J15, Z10

1 Introduction

Recent empirical evidence suggests that ethnic diversity has a negative impact on economic development and political stability. A high level of ethnic fractionalization is often associated with low levels of investment and worse institutional quality; in countries with high levels of ethnic polarization the probability of

*I am especially indebted to Francesco Caselli and Maitreesh Ghatak for their support. I also thank three anonymous referees, the editor Klaus F. Zimmermann, Oriana Bandiera, Erland Berg, Tim Besley, Matteo Cervellati, Raja Khali, Eliana La Ferrara, Andrea Prat and participants at NEUDC conference 2006, "Polarization and Conflict" conference 2006 and seminars at LSE. All errors are mine. University of Mannheim, Economics Department and SFB 884, L7, 3-5, 68131. Email: Paolo.Masella@mail.uni-mannheim.de.

civil conflict is higher.¹ "Nation building" (policies that promote attachment to nation over ethnic and regional identities) has been proposed as an integration and conflict reduction mechanism.² Miguel (2004) documents the implementation of nation building reforms in the newly independent East African nations in the 1960s and 1970s. He focuses on the economic development of two countries: Tanzania and Kenya. Despite their similar colonial institutional legacy, ethnic make up and geographical conditions, the leaders of these two countries have adopted very different policies, especially with regard to ethnic groups, over a wide range of dimensions. Nyerere (Tanzania) followed a "Pan-Africanist" nation building policy and a centralized economic policy. Miguel (2004) shows that nation-building allowed diverse communities in Tanzania to achieve better economic outcomes than diverse communities in the Kenyan regions.

The aim of the present work is to investigate the main determinants of national sentiment and, more importantly, the relationship between ethnic diversity and intensity of national feeling. Figures 1 and 2 cast some doubts on the conventional views. They plot ethnic diversity (in Figure 1 we use the index of ethnic fractionalization, in Figure 2 that of ethnic polarization)³ against the level of national identification of each country.⁴ Figures 1 and 2 suggest the absence of a negative relationship between ethnic heterogeneity and national sentiment.

[Insert Figures 1 and 2 here]

We use data from the World Values Survey (WVS) and estimate the likelihood that an individual identifies himself in national rather than in ethnic terms. Contrary to conventional wisdom, we find no evidence of less intense national sentiment in more ethnically fragmented countries or in minority groups. However, we find that national feelings in a minority are higher (lower) than in a majority depending on the ethnic diversity of a country. On the one hand, in more ethnically diverse countries, minorities show less intense national sen-

¹Mauro (1995) claims that ethno-linguistic diversity has a direct negative effect on the level of investment. Easterly and Levine (1997) find that a high level of ethnic fragmentation has a negative impact on economic growth. Montalvo and Reynal (2005) suggest that ethnic (and religious) polarization is one of the factors explaining economic development through its impact on the probability of civil wars. For a more accurate survey of the literature on the benefits and the costs of diversity see Alesina and La Ferrara (2005).

²See, among others, Deutch and Foltz (1963) and Tilly (1975).

³The ethnic fractionalization index can be interpreted as measuring the probability that two randomly selected individuals in a country belong to different ethnic groups. The purpose of the ethnic polarization index is, instead, to capture how far the distribution of the ethnic groups is from a bipolar distribution, which represents the highest level of polarization.

⁴National feeling at country level is measured as the proportion of individuals who choose to identify with their nation rather than with their ethnic group.

timent than majorities; on the other hand, in less diverse countries, the reverse is true. We find also that in larger groups national feelings are weaker and that individuals with higher incomes are less likely to feel an association with their ethnic group.

This paper is closely related to the growing literature on endogenous identity (see, among others, Akerlof and Kranton (2000), Bisin and Verdier (2000), Bisin et al. (2006) and Caselli and Coleman (2006)). In particular, the paper is connected to a recent literature that studies both the identity formation process and how identity can affect individual outcomes. Using the Afrobarometer surveys for nine Sub-Saharan African democracies, Bannon et al. (2004) estimate the likelihood that an individual identifies him/herself in ethnic terms rather than in terms of class or religion. They find that the salience of ethnicity is negatively related to ethnic diversity and claim that exposure to competition for jobs and political power are factors that predispose individuals to identify themselves in ethnic terms. The present paper focuses instead on national identity and attempts to study how the intensity of national sentiment differs depending on the characteristics of the ethnic group to which an individual belongs. Bisin et al. (2008), Manning and Roy (2009) and Constant et al. (2008) explore the importance of religion in shaping ethnic and national identity. Georgiadis and Manning (2008) provide evidence supporting the view that multicultural policies do indeed promote the integration and assimilation of immigrants in the UK; Clots-Figueras and Masella (2010) investigate how governments can influence individual identity through the education curricula.⁵ Charness et al. (2007) and Chen and Li (2009) conduct laboratory experiments to show how group identity affects individual behavior.

The rest of the paper is organized as follows. Section 2 describes the data and presents the empirical evidence. Section 3 discusses possible channels through which ethnic composition can affect the intensity of national feeling. Section 4 provides some concluding remarks.

⁵The relationship between education and identity is also explored by Aspachs et al. (2008). They study how students in the Basque Countries sort into schooling systems based on parents' identity.

2 Empirical evidence

2.1 Data

The data for this study come from the World Values Surveys (WVS) waves two (1990-1993), three (1995-1997) and four (1998-2000). This is a multi-country survey project that employs standardized questionnaires to investigate citizens' attitudes in a large set of countries.⁶ We focus especially on one of the questions in the survey (here, we provide the US example):

"Which of the following best describes you? 1 above all, I am an Hispanic American 2 above all, I am a Black American 3 above all, I am a White American 4 above all, I am an Asian American 5 I am an American first and a member of some ethnic group second"

We build a variable, *"national identity"*, which is equal to 1 if individuals answer *"I am an American first and a member of some ethnic group second"*, and 0 otherwise. It is constructed from the 33,904 responses to the question from individuals in 25 separate WVS rounds conducted in 21 countries: the United States, Canada, Spain, China, Brazil, Byelorussia, Latvia, Uruguay, Armenia, Azerbaijan, Macedonia, Georgia, Albania, Bosnia, Indonesia, Israel, Moldova, Bulgaria, Pakistan, Singapore and Jordan.

The World Values Survey involves face to face interviews where interviewers code the respondents' ethnic characteristics based on observation.⁷ We create a variable *"minority"* which is equal to 0 if the ethnic group to which the individual belongs is the largest in the country, and 1 otherwise.⁸ We include controls for individual income (each respondent is asked to choose from 10 income categories, net of transfers and taxes), individual education (*"primary"* is a dummy variable equal to 1 if primary is the highest level of education achieved by the respondent) and individual occupation. Unfortunately, income is coded differently in the survey of Bulgaria and in the third wave of the survey conducted in Macedonia;⁹ in those two cases we code "income" as missing. In the main

⁶Li (2010) uses World Values data and minority-majority categorizations to study the impact of social identities on tax attitudes.

⁷In the case of Canada we use a question about the language spoken at home to further distinguish between French Canadian and English Canadian.

⁸Jordan is an exception. The ruling ethnic group is the Jordanian group, which is actually slightly smaller than the Palestinian group, a historically discriminated group. In this case we classify Palestinian as the minority and Jordanian as the majority. The results are robust to the exclusion of Jordan from the dataset, however.

⁹In the case of Bulgaria there are 168 income categories rather than the standard 10; in the case of Macedonia (wave III) there are 88 categories. In the fourth wave in Macedonia income is coded in the standard way.

specification we include income as a control, but we also report results without income as control and, therefore, including Bulgaria and Macedonia (III wave) in the sample. The survey asks individuals about their main occupation. They are given 13 options to choose from. The omitted category is "agricultural worker".

We also include characteristics of the ethnic groups and of the countries as explanatory variables. The dataset in Fearon (2003) provides the size of the ethnic groups in the sample.¹⁰ As a proxy of the average income of the group, we use sample averages.

We include in our sample only those countries where there is perfect correspondence between (i) the ethnic groups identified by Fearon in his dataset (ii) the ethnic groups coded by the interviewer (iii) the ethnic groups included in the "identity" question.¹¹ To measure the ethnic diversity of a country we use two indices: the index of ethnic fractionalization and that of ethnic polarization. The former is quite common in the literature, while the index of ethnic polarization has been used only more recently.

The fractionalization index is defined as

$$Frac = 1 - \sum_{i=1}^N q_i^2$$

This indicator can be interpreted as measuring the probability that two randomly selected individuals in a country will belong to different ethnic groups. Therefore this index increases when the number of groups increases.

We also use the index **RQ** proposed by Reynal-Queirol

$$RQ = 4 \sum_{i=1}^N q_i^2 (1 - q_i)$$

where q_i is the size of the ethnic group i . The purpose of this index is to capture how far the distribution of the ethnic groups is from a bipolar distribution,

¹⁰Ethnic groups in Fearon's dataset are identified sometimes in terms of language and sometimes in terms of race. Although it would be interesting to separate the effects of race and language, the small size of the sample (number of countries included) does not allow this.

¹¹For instance, we omit countries such as South Africa where there is a discrepancy between the WVS and Fearon's dataset. Fearon's dataset codes 14 South African groups (Gname, Zulu, Xhosa, North Sotho, Tswana, Coloured, Afrikaner, South Sotho, English-Speaking, Tsonga, Swazi, Asian, Venda, Ndebele), the WVS only 4 groups (white, black, colored and Indian), which makes it impossible to interpret the coefficient of the interaction term between the variable "minority" and the country level of diversity (this coefficient is crucial for our purposes as explained in the next section).

which is the highest level of polarization.¹²

We use two weight variables provided by the survey. One is a national weight, which reflects the national distribution of key variables such as the urban-rural divide, education, demographics and economic activity, while the second variable assigns the same weight to every country in order to avoid a large sample bias in the pooled country study. Table 1 reports the share of respondents who identify themselves with their country for each of the 25 surveys in our sample. Table 2 presents descriptive statistics for the main variables used in the analysis.¹³

[Insert Tables 1 and 2 here]

2.2 Empirical strategy and results

2.2.1 Individual sources of national identification

We begin by regressing the variable "national identity" on a set of individual characteristics. Standard errors are always clustered at the country-time level. Column 1 of Table 3 reports our minimal specification. Respondents with higher incomes are more likely to choose nation over ethnic group. Being part of a minority group does not affect individual identity. Column 2 introduces age dummies and Column 3 includes country-year fixed effects. Column 4 includes controls for individual's occupation and education. The coefficient of the variable "primary" is negative and significant. This suggests that less educated respondents (with at most primary education) display weaker national feelings than better educated respondents.¹⁴ However, the coefficient is not very large. An individual educated to higher than primary level is (almost) 3 per cent more likely to identify himself with the nation than an individual with, at most, primary level education. It turns out that individual identity does not vary very strongly with occupation (the omitted category is "agricultural worker"). Individuals belonging to categories such as "professional" (teachers, lawyer...), "farmer" and to some extent "foreman" are more likely to be identified with the nation than "agricultural workers". The coefficient of "minority" is never significant in any of the four specifications, while the coefficient of income is always positive and significant.

[Insert Table 3 here]

¹²See Montalvo and Reynal (2005) for a detailed discussion.

¹³Descriptive statistics refer to the sample excluding Bulgaria and Macedonia (III wave).

¹⁴Similar findings are reported by Dustmann (1996). He finds that the level of education is positively correlated with the degree of assimilation of immigrants.

2.2.2 National identity and ethnic diversity

As a further step, we check if there is a relationship between identification and ethnic diversity. We regress identity on the set of individual controls included in our minimal specification (Table 3, Column 1), GDP per capita and ethnic diversity (in Table 4, Column 1 uses the index of ethnic fractionalization, Column 3 that of ethnic polarization). Columns 2 and 4 of Table 4 exclude "income" from the set of controls, therefore including Bulgaria and Macedonia (III wave) for the reasons discussed in Section 2.1. We find no evidence of a negative relationship between ethnic diversity and national identity; the results in the table show instead weak evidence of a positive relationship. However, the size of the sample is limited and the countries included in our study may not be a good substitute for a world-wide sample.

We turn now to the main specification. In the previous section, we show that the coefficient of the variable *minority* is not significant. Respondents from minority groups do not tend to be more or less identified with their nation than respondents from majority groups. In this section, we study whether the feelings of respondents from minority groups are influenced by the degree of ethnic diversity of the country in which they live. We introduce an interaction term between the variable "minority" and a country variable that measures ethnic diversity.

The specification of the pooled cross-country analysis is given by:

$$\begin{aligned} \textit{national identity}_{i,c,t} = & \alpha_{c,t} + \textit{income}_{i,c,t}\beta + \textit{minority}_{i,c,t}\gamma + \\ & + \textit{minority}_{i,c,t} * \textit{ethnic diversity}_c\delta + \varepsilon_{i,c,t} \end{aligned}$$

There are country-time fixed effects, which means that omitted country characteristics, correlated with ethnic polarization, are not affecting individual identity. However, there can be omitted country level variables having differential impacts on individual identity and correlated with right hand side variables (ethnic diversity in this case).

In Table 4, Columns 5 and 6 use the index of ethnic fractionalization, Columns 7 and 8 that of ethnic polarization; Columns 5 and 7 include income among the controls. We briefly describe the results reported in Table 4, Columns 5 to 8. The coefficient of the variable *minority* is positive while the sign of the coefficient of the interaction term is negative. Both coefficients are significant. In countries with average ethnic polarization, the identity choices of minority and that of the majority groups are no different. However, if we

increase the ethnic polarization index by one standard deviation, we find that individuals from ethnic minority groups are 10 percentage points less likely to identify with nation than those from the majority group, while if we decrease the ethnic polarization index by one standard deviation, we find that individuals from ethnic minority groups are 10 percentage points more likely to identify with nation than those from the majority group.¹⁵

[Insert Table 4 here]

Table 5 reports the results of several exercises performed to check the validity of the findings in Table 4, Columns 5-8. All the specifications include income among the control variables. In Row 1 the most fractionalized country in the sample is excluded, in Row 2 the least fractionalized country is excluded; in Row 3 the most polarized country is excluded, in Row 4 the least polarized country is excluded. In order to control for the possibility that we are capturing the effect of omitted country-variables (correlated with diversity) having differential impacts on individual identity, in Rows 5 and 6 our specifications include an interaction term between the variable "minority" and per capita GDP. In all six specifications the coefficient of the variable "minority" is positive and significant and the coefficient of the interaction term between ethnic diversity and "minority" is negative and significant.

[Insert Table 5 here]

2.2.3 Size of the ethnic group

Through a second set of regressions we study the relationship between national identity and the size of the ethnic group to which the individual belongs. In all the specifications income is included as one of the control variables. We divide the sample in two subsamples: one that includes only individuals belonging to a minority group and one that includes only individuals belonging to the majority group. We regress identity on the usual set of individual characteristics and on the size of the group to which the respondent belongs. Table 6, Columns 2 and 5 include the sample mean income for each ethnic group. We can only include country fixed effects when we use the subsample including only individuals from minority groups. The coefficient of the variable measuring the size of the ethnic group to which the respondent belongs is always negative and significant. As before, no causality can be claimed; omitted group characteristics (and country characteristics when country fixed effects are not included), correlated with the

¹⁵The results reported in Table 4 are robust to the exclusion from the dataset of the 3 countries (Bosnia, Indonesia and Jordan) where the ethnic majority represents less than 50% of the population.

size of the group, may affect individual identity.

[Insert Table 6 here]

2.3 Measurement issues and other concerns

There are several limitations to our analysis. First, respondents' answers are context specific. We try to control for this by using what we know about the survey context (particularly where and when the survey was conducted). However, there are aspects for which we cannot control, such as the proximity of ethnic festivals.

Secondly, there may be measurement errors, which are particularly important in attitude surveys. Respondents from societies where social norms prevent any open talk about ethnicity may be less likely to declare ethnic identification.¹⁶ In countries where ethnicity is defined by race, increased awareness of racism might affect the willingness of a white person to say "I am first and foremost white" and make them more likely to identify with the nation, which might induce an upward bias on measured national identity. However, the question analyzed is only one out of 220 questions in the standard World Values Survey, and the only one that refers explicitly to ethnicity. Also, the inclusion of country fixed effects partly deals with this concern.

Thirdly, it is not clear whether the countries in our study are a good substitute for a world-wide sample. Our findings should be interpreted with the caveat that they may not be representative of the whole world.

We found that income is positively correlated with the likelihood that an individual identifies him/herself in national rather than ethnic terms. The size of the group is negatively correlated with the intensity of national feelings. Those relationships, however, cannot be claimed to be causal since omitted variables correlated with size or income may affect national identity. In the case of the main specification, the inclusion of country fixed effects moderates the omitted variables problem. However, there may still be omitted country-variables that have differential impacts on individual identity and which are correlated with the right hand side variables (ethnic diversity in this case).

¹⁶The survey was conducted by interviewers who were not affiliated to any political party or government, making it likely that the survey was not perceived as related to a national institution.

3 Discussion of the Channels

In this section we discuss the channels through which a country's ethnic make-up might affect the strength of national feeling.

In very diverse societies there is likely to be more economic, political and cultural discrimination towards minorities, which may be directly detrimental to the degree of integration of minorities or indirectly through the reduction of their income and economic opportunities.¹⁷ This could explain why, in very diverse societies, individuals from minorities groups tend to be less nationalistic than individuals from majorities.¹⁸

A governance outcome commonly associated with severely divided societies is patronage. Patronage refers to the system of granting benefits to members of a particular ethnic group (the one in power) while discriminating against other ethnic groups. Examples of groups advantaged by their political leaders are the Northern groups in Nigeria and Uganda, and the Tutsis in Burundi. Alesina, Baqir and Easterly (1998) provide more formal evidence showing that public employment is significantly higher in US cities where ethnic fragmentation is higher. They interpret public employment as a subsidy to ethnically defined interest groups.

There is substantial evidence also that discrimination reduces cultural integration and the establishment of national identity among minorities. Using data from the 2007 UK Citizenship Survey, Georgiadis and Manning (2009) find that members of minority groups who feel well treated tend to identify more strongly with the UK, while those, who perceive to be treated badly by a variety of public sector organizations or who feel that their ethnic groups are discriminated against by British society, tend to report weaker national feelings.¹⁹ Bisin et al. (2008) using a different data source (UK Fourth National Survey of Ethnic Minorities) obtained similar findings; in the UK, episodes of harassment and discrimination in the work place for reasons of race or color, or religious or cultural beliefs seem to be strongly positively correlated with the intensity of religious feeling among minorities. Discrimination and negative public attitudes are reported as the most important barriers to the integration of immigrants and minorities by Constant et al. (2009) based on their analysis of the new IZA

¹⁷This is consistent with the positive association found between income and national feeling.

¹⁸Krueger and Pischke (1997) and Dustmann and Preston (2001) suggest that it is likely that high concentrations of ethnic minorities promoted more hostile attitudes against minorities in Germany and England.

¹⁹Bauer et al. (2000) suggest that attitudes of the native population towards immigration is likely to be related to whether immigrants are selected according to the needs of the labor markets.

Expert Opinion Survey conducted in 2007 among expert stakeholders in the EU-27. Using data from the Afrobarometer, Bannon et al. (2004) find that race is most salient in countries with recent histories of racial discrimination (South Africa, Zimbabwe, and Namibia). Conflicts between cultures, however, are not inevitable; multicultural policies might improve the level of assimilation of immigrants and minorities. Citizenship rights are correlated with national feelings in the UK (Manning and Roy, 2009) and the US; Avitabile et al. (2010) analyze a provision of the 2000 German nationality reform and show that introducing elements of birthright citizenship in Germany increased the level of integration of immigrants measured by their propensity to engage in social contacts with Germans and to use the German language.

As discussed above, the results in Table 4 imply that in countries with low ethnic diversity individuals from minorities tend to show stronger national sentiment than individuals from majorities. This might be explained by a direct negative effect of group size on the intensity of national identity. Each ethnic group has its own social norms²⁰ and deviations from them might involve some psychological costs.²¹ It might be that in order to overcome the shocks related to the distance from a particular ethnic group, an individual develops stronger national feeling. In other words, nationalist sentiment may represent a form of compensation for the abandonment of the ethnic convention. These costs are likely to be higher the more widespread the norm in the population (and the larger the ethnic group). Then, individuals from smaller ethnic groups will deviate more often from their groups' norms and will develop stronger national feelings.²² This is in line with the assumption used to construct the polarization index that identification with one's own group is a positive function of its size²³ and, more importantly, is confirmed by the results in Table 6. Group size seems to be negatively associated with the probability that respondents choose nation over ethnic identity.

²⁰Members of an ethnic group might be expected to consume a specific ethnic good as defined by Chiswick (2009), respect a specific dress code (e.g. the Islamic veil) or a particular diet, as discussed by Epstein (2006). Norms can involve the use of the local language at home (Lazear, 1999) or participation in rituals and festivals (Kuran, 1998).

²¹This is consistent with a broad class of conformity models (see, among others, Akerlof, 1980 and Kandel and Lazear, 1992).

²²Similar assumptions are made by Bisin et al. (2006) and Akerlof (1980). Kandel and Lazear (1992) discuss the relationship between firm size and peer pressure; they suggest that the level of monitoring may increase with the size of the firm and also that, if more workers observe an individual, the sanctions imposed might be greater.

²³See Esteban and Ray (1994).

4 Conclusions

This paper investigated the relationship between ethnic diversity and individual identity. Using data from the World Values Surveys, we provide several empirical facts. There is no evidence of a lower intensity of national sentiment in more ethnically fragmented countries or in minority groups. National feelings in a minority can be higher or lower than in a majority, depending on the degree of ethnic diversity of the country. On the one hand, in very ethnically diverse countries, minorities have weaker national sentiments than majorities; on the other hand, in countries with low ethnic diversity, the reverse is true.

It is commonly believed that creation of a shared identity among the citizens of a country might be beneficial from an economic point view. There is substantial evidence that more pro-social behavior is observed among groups of individuals who perceive themselves as sharing the same culture and identity. This is likely to be important in countries with high levels of diversity since empirical studies show the existence of a strong negative correlation between ethnic diversity and both economic development and political stability. One of the aims of this paper was to identify which countries should pursue more aggressive nation building programs and whether there are certain groups that should be targeted. Minority groups in ethnically diverse countries and majority groups in homogeneous societies seem to show weaker intensity of national identity. In heterogeneous societies, where economic outcomes are worse, our findings seem to suggest the need for policies that are able to stimulate feelings of national identity among minorities, such as the enforcement of a common language for education and administration in countries where ethnic groups tend to be differentiated along language lines (this applied to Tanzania in the mid-1960s; for an extensive discussion see Miguel (2004)).

5 Data Appendix

5.1 Definition of variables

-*National identity*: dummy equal to 1, if the respondent says "*I am an American first and a member of some ethnic group second*" (US example). Source: WVS

-*Minority*: dummy equal to 0, if the ethnic group to which the individual belongs is the largest in the country; otherwise it is equal to 1. Each interviewer has been asked to code the ethnic group of each individual in the sample. Source: WVS

-*Age*: age of the respondent. Source: WVS

-*Female*: dummy equal to 1, if the respondent is female. Source: WVS

-*Primary*: dummy equal to 1, if the respondent has, at most, primary school education. Source: WVS

-*Married*: dummy equal to 1, if respondent is married. Source: WVS

-*Income*: Proxy for individual income stream (not disposable). There are 10 income categories (net of transfers and taxes); they have been coded by deciles for each country, 1=lowest decile; 10=highest decile. Source: WVS

-*Employer>10*: dummy equal to 1, if the respondent is employer/manager of an establishment with 10 or more employees. Source: WVS

-*Employer<10*: dummy equal to 1, if the respondent is employer/manager of an establishment with less than 10 employees. Source: WVS

-*Professional*: dummy equal to 1, if the respondent is a professional worker, lawyer, accountant, teacher, etc Source: WVS

-*Supervisor*: dummy equal to 1, if the respondent is a supervisory, non-manual office worker Source: WVS

-*Office*: dummy equal to 1, if the respondent is a non-supervisory, non-manual office worker. Source: WVS

-*Foreman*: dummy equal to 1, if the respondent is a foreman and supervisor Source: WVS

-*Skilled*: dummy equal to 1, if the respondent is a skilled manual worker Source: WVS

-*Semi-skilled*: dummy equal to 1, if the respondent is a semi-skilled manual worker Source: WVS

-*Unskilled*: dummy equal to 1, if the respondent is an unskilled manual worker Source: WVS

-*Farmer*: dummy equal to 1, if the respondent has his/her farm Source: WVS

-*Armed forces*: dummy equal to 1, if the respondent is a member of armed forces. Source: WVS

-*Never job*: dummy equal to 1, if the respondent never had a job Source: WVS

-*Size*: size of ethnic group to which the individual belongs. Source: Fearon's dataset

-*Etfra*: country index of ethnic fractionalization. Source: Fearon's dataset

-*Etpol*: country index of ethnic polarization. Source: Fearon's dataset

5.2 Countries and ethnic groups

Albania (wave IV): Albanian, Greek, other; **Armenia** (wave III): Armenian, Russian, other; **Azerbaijan** (wave III): Azerbaijani, Russian, other; **Belarus** (wave III): Belarusian, Russian, Polish, Ukrainian, other; **Bosnia** (wave IV): Bosniak, Serbs, Croats, other; **Brazil** (wave III): white, mulatto, black, other; **Bulgaria** (wave III): Bulgarian, roma, Turkish, other; **Canada** (wave IV): English Canadian, French Canadian, black, South Asian, Chinese, East Asian, indigenous peoples, other; **China** (wave II): Han, other; **Georgia** (wave III): Georgian, other; **Indonesia** (wave IV): Javanese, Malay, Chinese, Sundanese, other; **Israele** (wave IV): Jewish, Arabic; **Jordan** (wave IV): Jordanian, Palestinian, other; **Latvia** (wave III): Latvian, Russian, Belarusian, Ukrainian, Polish, other; **Macedonia** (waves III-IV): Macedonian, Albanian, Turkish, roma, Serb, other; **Moldova** (wave III): Moldovian, Slavs, Bulgarian, Gaugas; **Pakistan** (wave IV): Punjabi, Baluchi, Sindhi, Urdu speaking, other; **Singapore** (wave IV): Malay, Indian, south Asian, Chinese, other; **Spain** (waves III-IV): Castellano, Catalan, Basque, Galician, other; **Uruguay** (wave IV): white, black, other; **US** (waves II-III-IV): white, black, Asian, Hispanic, other;

References

- Akerlof GA (1980) A Theory of Social Custom of which Unemployment May be One Consequence. *Quarterly Journal of Economics* 94, 749-775
- Akerlof GA, Kranton RE (2000) Economics and identity. *Quarterly Journal of Economics* 115, 715-753
- Alesina A, Baqir R, Easterly W (2000) Redistributive Public Employment. *Journal of Urban Economics* 48, 219-241
- Alesina A, La Ferrara E (2005) Ethnic diversity and economic performance. *Journal of Economic Literature*, vol. 43, pp. 762–800
- Aspachs-Bracons O, Clots-Figueras I, Masella P (2008) Compulsory Language Educational Policies and Identity Formation. *Journal of the European Economic Association*, 6, 434-444
- Avitabile C, Clots-Figueras I, Masella P (2010) The Effect of Birthright Citizenship on Parental Integration Outcomes. CSEF Working Paper n. 246

- Bannon A, Miguel E, D Posner (2004) Sources of ethnic identification in Africa. Afrobarometer Working Paper No. 44
- Bauer T, Lofstrom M, Zimmermann KF (2000) Immigration Policy, Assimilation of Immigrants and Natives' Sentiments towards Immigrants: Evidence from 12 OECD-Countries. *Swedish Economic Policy Review*, (2), 11-53
- Bisin A, Verdier T (2000) The economic of cultural transmission and the dynamics of preferences. *Journal of Economic Theory* 97, 298-319
- Bisin A, Patacchini E, Verdier T, Zenou Y (2006) 'Bend it like Beckham': identity, socialization and assimilation. CEPR Discussion paper No 5662
- Bisin A, Patacchini E, Verdier T, Zenou Y (2008) Are Muslim Immigrants Different in Terms of Cultural Integration?. *Journal of the European Economic Association*, 6, 445-456
- Caselli F, Coleman W (2006) On the theory of ethnic conflict. NBER Working Paper No. W12125
- Charness G, Rigotti L, Rustichini A (2007) Individual Behavior and Group Membership. *American Economic Review*, 97, September, 1340-1352
- Chen Y, Li SX (2009) Group Identity and Social Preferences. *American Economic Review*, 99(1): 431-57
- Chiswick B, Miller P (1999) Language skills and earnings among legalized aliens. *Journal of Population Economics*, vol. 12(1), p. 63-89
- Chiswick C (2009) The economic determinants of ethnic assimilation. *Journal of Population Economics*, vol. 22(4), pages 859-880
- Clots-Figueras I, Masella P (2010) Education, Language and Identity. Mimeo
- Constant A, Gataullina L, Zimmermann KF (2008) Ethnosizing Immigrants. *Journal of Economic Behavior and Organization*, 69, 274-287
- Constant A, Kahanec M, Zimmermann KF (2009) Attitudes towards immigrants, other integration barriers, and their veracity, *International Journal of Manpower*, Emerald Group Publishing, vol. 30(1/2), pages 5-14, May
- Deutch K, Foltz W (1963) Nation building. Atherton press, New York
- Dustmann C, Preston I (2001) Attitudes to Ethnic Minorities, Ethnic Context, and Location Decisions. *Economic Journal*, 111(470), 353-373

- Dustmann C (1996) The social assimilation of immigrants. *Journal of Population Economics* , 9 (1) , pp. 37-54
- Easterly W, Levine R (1997) Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics* 112(4), 1203-1250
- Epstein GS (2006) Migrants, Ethnicity and Employment. Unpublished manuscript
- Esteban J, Ray D (1994) On the measurement of polarization. *Econometrica*, 62, 4, 819-8510
- European Values Study Group and World Values Survey Association. EUROPEAN AND WORLD VALUES SURVEYS INTEGRATED DATA FILE
- Fearon J (2003) Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, 8(2), 195-222
- Georgiadis A, Manning A (2008) One Nation Under a Groove? Multiculturalism in Britain. Centre for Economic Performance Discussion Paper, London School of Economics, November
- Kandel E and Lazear EP (1992) Peer Pressure and Partnerships. *Journal of Political Economy* 100, 801-817
- Kreuger A, Pischke JS (1997) A Statistical Analysis of Crime Against Foreigners in Unified Germany, *Journal of Human Resources*, 32 (1), 182-209
- Kuran T (1998) Ethnic Norms and Their Transformation Through Reputational Cascades. *Journal of Legal Studies* 27(2): 623-659
- Li XS (2010) Social Identities, Ethnic Diversity, and Tax Morale. *Public Finance Review*, Vol. 38, No. 2, March 2010: 146-177
- Lazear E (1999) Culture and Language. *Journal of Political Economy*, 107(S6), S95-29
- Mauro P (1995) Corruption and Growth. *Quarterly Journal of Economics* CX: 681-712
- Manning A., Roy S. (2009) Culture Clash or Culture Club? National Identity in Britain. *The Economic Journal*, Vol. 120, No. 542, pp. F72-F100, February 2009

- Miguel E (2004) Tribe or nation? Nation-Building and Public goods in Kenya versus Tanzania. *World Politics*, 56 (3), 327-362
- Montalvo J, Reynal M (2005) Ethnic diversity and economic development. *Journal of Development economics*, 76, 293-323
- Tilly C (1975) *The formation of national States in Western Europe*. Princeton University Press, Princeton

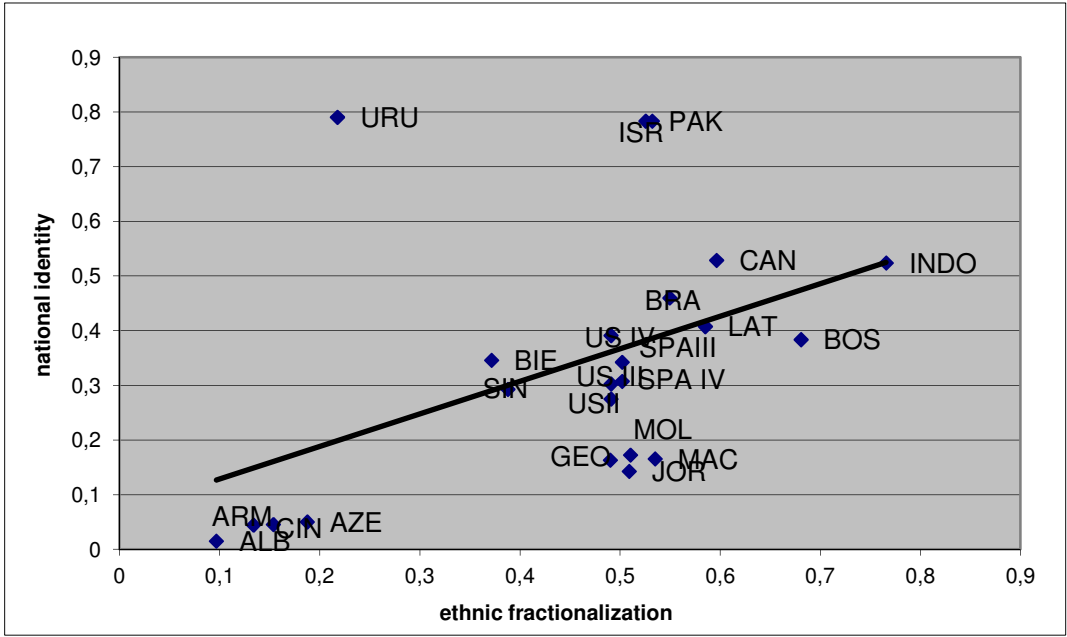


Figure 1: National Identity and Ethnic Fractionalization

Note: Country level National Identity is measured as the fraction of individuals in each country that chooses to identify with the nation rather than with its ethnic group.

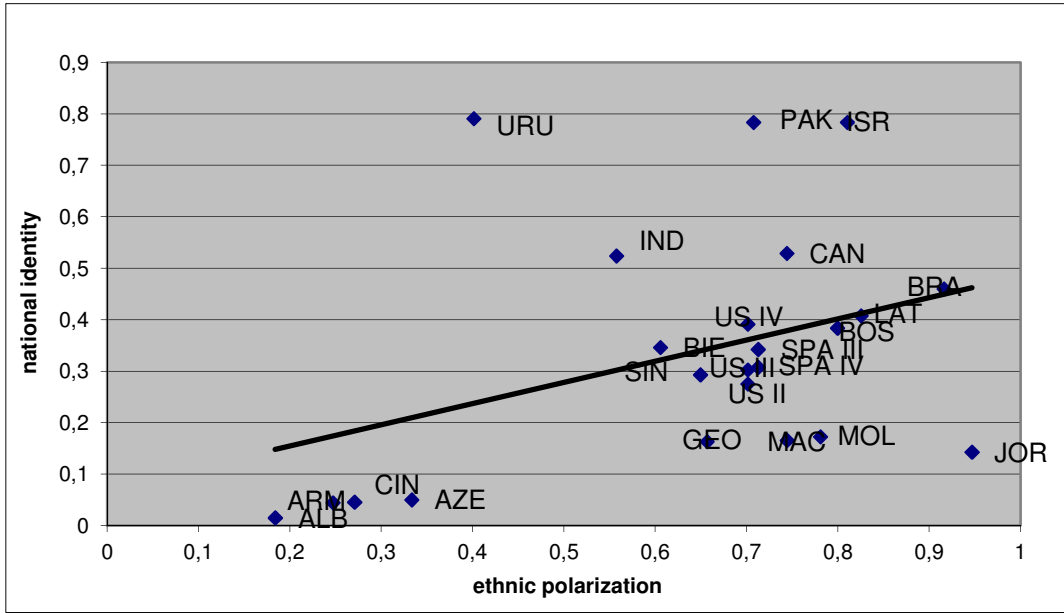


Figure 2: National Identity and Ethnic Polarization

Note: Country level National Identity is measured as the fraction of individuals in each country that chooses to identify with the nation rather than with its ethnic group.

Table 1: National identity by country

National Identity		National Identity	
Albania (w. IV)	0,015	Latvia (w. III)	0,4077
Armenia (w. III)	0,0444	Macedonia (w. III)	0,0145
Azerbaijan (w.III)	0,0501	Macedonia (w. IV)	0,1655
Belarus (w. III)	0,3459	Moldova (w. III)	0,1726
Bosnia (w.IV)	0,3837	Pakistan (w. IV)	0,7835
Brazil (w. III)	0,4599	Singapore (w. IV)	0,2929
Bulgaria (w.III)	0,0541	Spain (w. IV)	0,3075
Canada (w. IV)	0,5288	Spain (w. III)	0,3424
China (w. II)	0,0454	Uruguay (w.IV)	0,7906
Georgia (w.III)	0,1633	US (w. IV)	0,3912
Indonesia (w. IV)	0,524	US (w. III)	0,3025
Israele (w.IV)	0,7835	US (w. II)	0,2752
Jordan (w.IV)	0,1427		

Note: Country level National Identity is measured as the fraction of individuals in each country that chooses to identify with the nation rather than with its ethnic group.

Table 2: Descriptive Statistics

	Mean	St. dev.	obs
Age	41	16,39	32424
Female	0,51	0,49	32457
Married	0,57	0,49	32484
Income	4,41	2,36	29673
manager>10	0.02	0.16	30460
manager<10	0.04	0.2	30460
professional	0.13	0.34	30460
supervisor	0.06	0.24	30460
nonmanual	0.09	0.28	30460
foreman	0.03	0.18	30460
skilled	0.17	0.37	30460
semiskilled	0.07	0.26	30460
unskilled	0.09	0.28	30460
farmer	.02	0.15	30460
armed forces	0.01	0.13	30460
never worked	0.17	0.38	30460
primary	0.23	0.42	29636
Av. group income	4,49	1,2	95
minority	0,25	0,43	32498
etfra	0,44	0,17	23
size	0,23	0,28	95
etpol	0,63	0,21	23
national identity	0,32	0,46	32087

Note: The table reports the mean values of variables in the sample with standard deviations in parentheses. The Descriptive Statistics refer to the sample without Bulgaria and Macedonia (III wave). See the text and the Data Appendix for details and definitions.

Table 3 Sources of national identification: individual characteristics

	national identity			
Age	0.002**			
	(0.00)			
Female	0.004	0.002	-0.003	-0.003
	(0.011)	(0.011)	(0.008)	(0.009)
married	-0.041*	-0.06**	0.005	0.002
	(0.23)	(0.023)	(0.009)	(0.01)
income	0.02***	0.019***	0.007*	0.005*
	(0.06)	(0.006)	(0.002)	(0.003)
minority	0.003	0.001	-0.049	-0.052
	(0.06)	(0.06)	(-0.048)	(-0.048)
primary(at most)				-0.029*
				(0.016)
man>10				0.002
				(0.027)
man<10				0.03
				(0.029)
professional				0.055**
				(0.024)
supervisor				0.004
				(0.022)
nonmanual				0.002
				(0.021)
foreman				0.04*
				(0.021)
skilled				0.004
				(0.019)
semiskilled				0.004
				(0.015)
unskilled				0.006
				(0.023)
farmer				0.045**
				(0.018)
armforc				0.036
				(0.025)
neverworked				0.008
				(0.02)
Age f.e.	no	yes	yes	yes
country-time f.e.	no	no	yes	yes
observations	29230	29230	29230	25132
country-time obs	23	23	23	20
R squared	0,015	0,022	0,217	0,229

Note: The dependent variable is a dummy that is equal to 1 if the respondent identifies with the nation rather than the ethnic group. *Minority* is a dummy variable that is equal to 1 if the individual belongs to a minority group. The variable *primary* is a dummy equal to 1 if the respondent has a low level of education (at most primary). *Income* is our proxy for the individual's income stream. Column 4 includes controls for respondent's profession. The US (II wave) and China samples do not provide information on respondent's education; the Israel sample does not provide information on respondent's profession. Standard errors clustered at country-time level are reported in parentheses. * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 4 Sources of national identification: Diversity

	National identity							
income	0.014** (0.005)		0.019*** (0.005)		0.007* (0.003)		0.007* (0.003)	
minority	-0.059 (-0.051)	-0.062 (0.048)	-0.036 (-0.05)	-0.04 (0.049)	0.344** (0.144)	0.349** (0.137)	0.328* (0.165)	0.382** (0.149)
etfra	0.55** (0.254)	0.579** (0.243)						
etfra*minority					-0.765** (0.304)	-0.779** (0.298)		
etpol			0.357* (0.208)	0.364* (0.192)				
etpol*minority							-0.531** (0.254)	-0.607** (0.236)
observations	29230	33904	29230	33904	29230	33904	29230	33904
country*time F.E.	no	no	no	no	yes	yes	yes	yes
country*time obs	23	25	23	25	23	25	23	25
R squared	0.06	0.049	0.045	0.0318	0.221	0.236	0.218	0.234

Note: The dependent variable is a dummy that is equal to 1 if the respondent identifies with the nation rather than the ethnic group. *Minority* is a dummy variable that is equal to 1 if the individual belongs to a minority group. The variables *etfra* and *etpol* represent the level of ethnic fractionalization and polarization in the country of residence of the respondent. *Income* is our proxy for the individual's income stream. In Columns 1-4 we include per capita GDP (at the time of the survey) as a control variable. All specifications control for gender, marital status and age (linearly). Standard errors clustered at the country-time level are reported between parentheses. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 5 Sources of national identification: Diversity (Robustness)

		minority	etfra*minority	etpol*minority
Row 1	Most Fract.	0.454***	-1.014***	
	excluded	(0.152)	(0.352)	
Row 2	Least Fract.	0.33**	-0.74**	
	excluded	(0.151)	(0.316)	
Row 3	Most Pol.	0.328*		-0.531**
	excluded	(0.165)		(0.254)
Row 4	Least Pol.	0.36*		-0.582*
	excluded	(0.205)		(0.331)
Row 5	GDP*minority	0.486***	-0.825***	
		(0.131)	(0.257)	
Row 6	GDP*minority	0.417**		-0.514**
		(0.155)		(0.229)

Note: The variables *etfra* and *etpol* represent the level of ethnic fractionalization and polarization in the country of residence of the respondent. Specifications in Rows 1, 2 and 5 use the specification in Table 4, Column 5 as a benchmark. Row 1 excludes the most fractionalized country, Row 2 excludes the least fractionalized, Row 5 includes an interaction term between per capita GDP and the variable "minority". Specifications in Rows 3, 4 and 6 use the Specification of Column 7 of Table 4 as a benchmark. Row 1 excludes the most polarized country, Row 2 excludes the least polarized, Row 6 includes an interaction term between per capita GDP and the variable "minority" .

Table 6: Sources of national identification: size of the group

	National identity				
av.group inc.		0.065 (0.041)			0.069* (0.35)
size	-0.674* (0.33)	-0.669** (0.32)	-0.695*** (0.145)	-0.765** (0.304)	-0.637** (0.283)
country*time F.E.	no	no	no	yes	yes
country*time obs	23	23	23	23	23
total obs.	21479	23523	7600	7600	7600
Sample	only maj	only maj	only min	only min	only min
R squared	0.063	0.079	0.066	0.18	0.183

Note: The dependent variable is a dummy that is equal to 1 if the respondent identifies with the nation rather than the ethnic group. The variable *size* measures the size of the ethnic group to which the respondent belongs to. *Average group income* measures the sample average of the incomes of the components of the ethnic group to which the respondent belongs. Specifications 1-2 include only respondents belonging to the majority group in the country of residence. Specifications 3, 4 and 5 include only respondents who do not belong to the majority group in the country of residence. All specifications control for individual income, gender, marital status and age (linearly). Standard errors clustered at group level are reported in parentheses. *Significant at 10%, ** significant at 5%, *** significant at 1%.