

Online Appendix

Intergroup Contact and Nation Building: Evidence from Military Service in Spain

Julio Cáceres-Delpiano* Antoni-Italo De Moragas†
Gabriel Facchini‡ Ignacio González§

September 23, 2021

*Department of Economics, Universidad Carlos III de Madrid. jcaceres@eco.uc3m.es

†Department of Economics, CUNEF. antoni.demoragas@cunef.edu

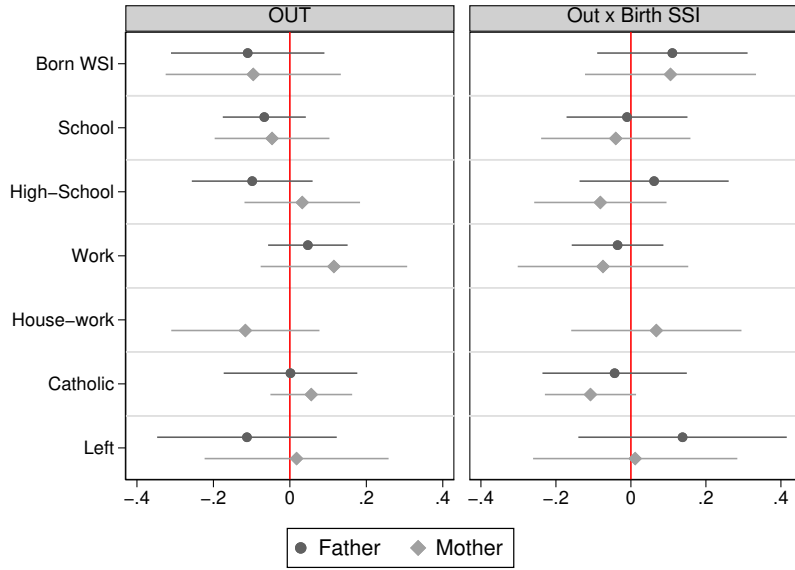
‡Department of Applied Economics, Universitat Autònoma de Barcelona. gabriel.facchini@uab.cat

§Department of Economics, American University. ignaciog@american.edu

A Balancing on pre-treatment observables

The nature of the lottery guarantees that being assigned to military service in another region is as good as random, conditional on the respondent's region of birth. Nevertheless, in this section, we show that these two groups of individuals do not differ in their pre-treatment observable characteristics. Given that all the respondent characteristics are observed several years after the treatment and could potentially be affected by it, the only observable characteristics that we are confident were not affected by the treatment are the parents' characteristics. We believe that the most relevant parental characteristic that could correlate with the respondent's national identity and his vote is the father's and mother's birth region.¹ In Figure A.1 we show that none of the pre-treatment characteristics, including the region of birth of the parents, is statistically different from zero between the respondents assigned to military service in their region and the respondents assigned to military service in another region.

Figure A.1: Balance of pre-treatment observables



Notes: The figure represents the coefficients and 95% confidence intervals from separate regressions of each predetermined variable by using the empirical specification of equation 2. All regressions include fixed effects for month of birth and fixed effects for year of birth interacted with region of birth fixed effects. Standard errors are clustered at the year of birth-province level. All variables regarding parents' characteristics are defined at the time that the respondent was 16 years old. Born WSI is a dummy equal to one if the parent was born in a region with a weak Spanish identity. School is a dummy equal to one if the parent attended school. High-school is a dummy equal to one if the parent finished high school. Work is a dummy equal to one if the parent was employed. House-Work is a dummy equal to one if the parent worked at home unpaid. Catholic is a dummy equal to one if the parent was Catholic. Left is a dummy equal to one if the respondent considered his parent to be more left than right (by using a ten-point scale and taking five or more).

¹Our survey does not include the national identity of the parents, and in any case, it would not be a pre-treatment observable since it could also be affected by the post-treatment contact with our respondent.

B Definition of Spanish Identity

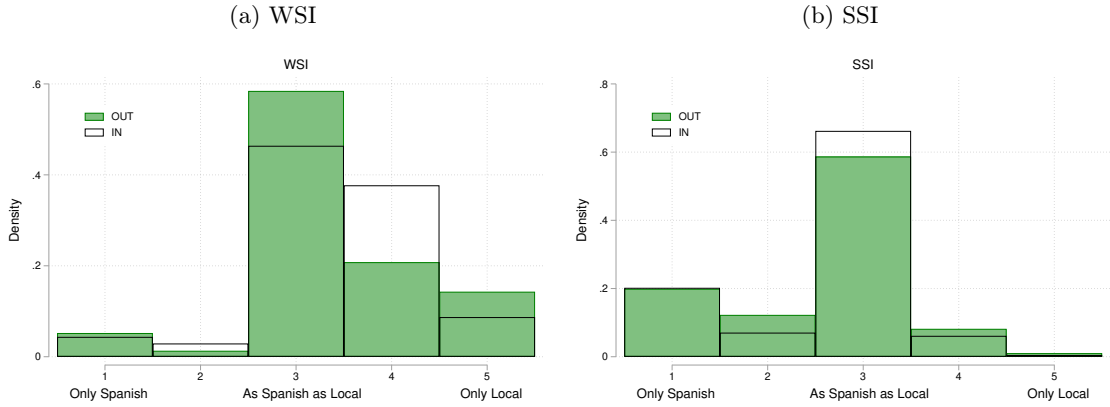
We made our decision of grouping the first 3 answers to the 5-item Likert-Scale question in one group as one (“only Spanish”, “more Spanish than local”, “as Spanish as local”), and the other two as zero (“more local than Spanish”, “only local”) based on previous literature and observed data.

Previous studies have highlighted the non-linearity of our dependent variable (the so-called Linz-Moreno question). In particular, Guinjoan and Rodon (2015) compare answers to the Linz-Moreno question with answers in scales of Catalan and Spanish nationalism. They show a nonlinear link between the values on the Spanish nationalism scale and the Linz-Moreno scale. More precisely, their empirical analysis shows that individuals with pro-Spanish identity positions tend to choose the dual-identity category (“As Catalan as Spanish”) of the Linz-Moreno question. In other words, in WSI regions such as Catalonia, variability between categories 1 to 3 in the scale fails to properly capture the intensity of self-declared Spanish identity. Therefore, the distinction between these categories cannot provide any information about the impact of intergroup contact on national identity. Further evidence of the non-linearity of the Linz-Moreno question is found in Rodon and Guinjoan (2018). In their paper, they present the probability of supporting independence for each of the five categories of our dependent variable and they find that a huge increase from category 3 to category 4 and very little variability between categories 1, 2 and 3. This suggests that the margin between 3 and 4 is where the Linz-Moreno question captures meaningful changes in national identity. This is the variation that we exploit to construct the dummy variable of Spanish identity in our paper.

We observe a similar behaviour in our data. Figure A.2 shows the histogram of the Likert-Scale variable for individuals from WSI (left) and SSI (right) regions by their IN or OUT status based on the lottery. For individuals from WSI regions there is a mass shift from category 4 (“more local than Spanish”) to category 3 (“as Spanish as local”). Instead, there is no such mass shift for individuals from SSI regions. Furthermore, there doesn’t seem to be other shifts in mass for WSI individuals (except for a small increase in category 5).

Finally, note that there is a very small number of observations in categories 1 and 2 (less than 7% of the sample from WSI regions report an identity in those categories). In fact, categories 3 and 4 account for 81,51% of the sample in WSI regions. Therefore it is reasonable to expect that any effect can be driven by these categories.

Figure A.2: Spanish Identity - 5-item Likert Scale



Notes: Subfigure A.2a plots the histogram of the original 5 item Likert Scale for individuals from WSi regions and Subfigure A.2b plots the histogram of the original 5 item Likert Scale for individuals from SSI regions. In each subfigure we distinguish between individuals assigned to do their military service in their region (IN) and individuals assigned to do their military service in another region (OUT).

C Definition of the treatment variable (OUT)

Our decision to define the treatment between individuals with positive probability of migrating and 0 probability is motivated by the particular form of the distribution of the probability of being assigned to do the military service in another region (see Figure A.3). We can observe that the distribution of probabilities is concentrated in the two extremes. In fact, more than 50% of the observations have a probability of zero. For this reason, we considered sensible to to construct a dummy variable equal to one for all case with non-zero probability. In that way, we have two groups with a similar number of observations.

Figure A.3: Histogram Probability OUT

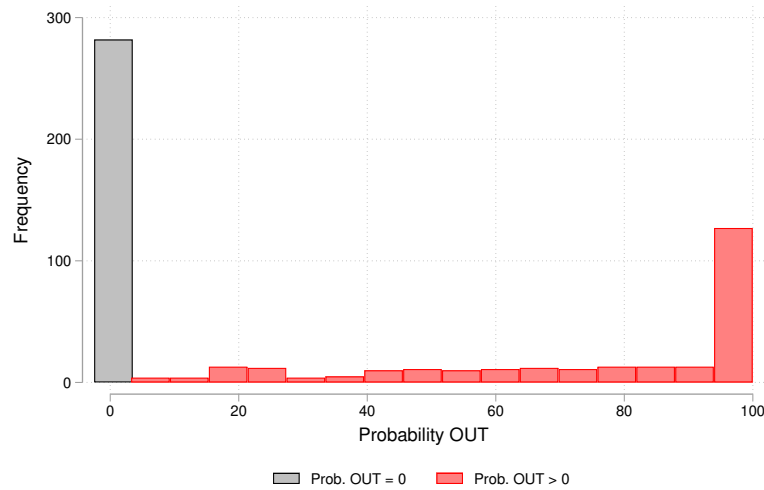
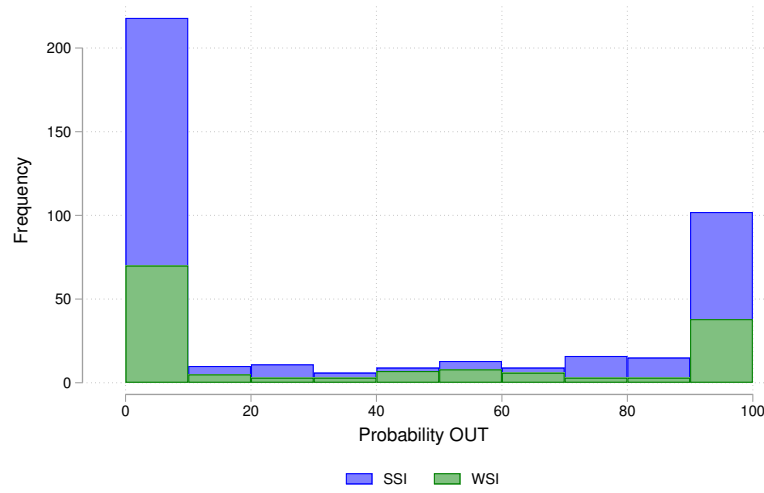
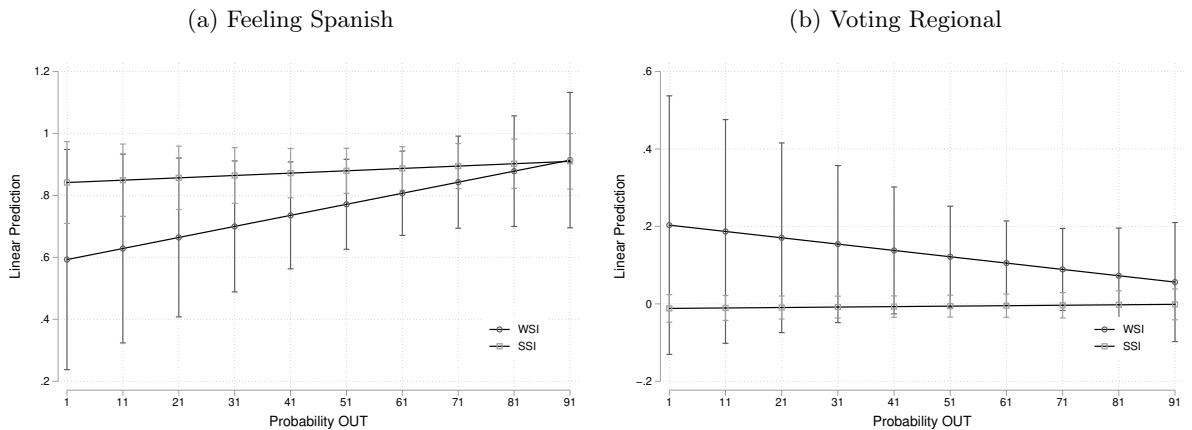


Figure A.4: Histogram Probability OUT - By Spanish Identity



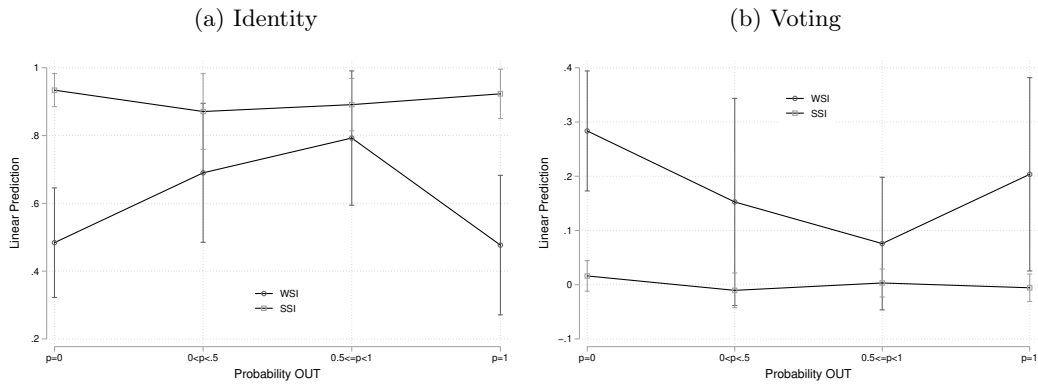
We have performed a robustness check where we exclude the two mass concentration points (0 and 100), and keep only individuals with a probability between 0 and 100. This significantly reduces the sample, but it helps use check whether the mass points can be driving our results. This is not the case. We run our main specification using only this restricted sample, and results are qualitatively the same. Figure A.5 shows the estimated probabilities at different values of our migration probability for our two main outcomes: identity (on the left) and voting (on the right). We can observe that individuals from SSI regions observe a similar probability of feeling at least as Spanish as local, irrespective of the probability of migrating. However, for individuals from WSI regions, this probability is increasing in the probability of migrating. A similar pattern emerges for voting for regional parties, only that the slope is negative, meaning that individuals from WSI regions are less likely to vote for a regional party the larger their probability of migrating to do the military service. One caveat of this exercise though is that, given the small sample size (130 observations) the standard errors are quite big.

Figure A.5: Main regression using only probabilities above 0 and below 1



Finally, we define the treatment as a categorical variable with 4 values: 0, if probability equals 0, 1, if probability is above zero but below 0.5, 2 if probability is above or equal 0.5 but below 1, and 3 if the probability equals 1. By using this variable, we allow for non-linear effects in different parts of the distribution of probabilities. We repeat our main specification using this categorical variable for national identity and voting for a regional party and report the predicted probabilities in Figures A.6a and A.6b, respectively. As expected, changes in the probability of being sent to another region have no effect on the probability of feeling more Spanish or voting for regional parties for individuals from SSI regions. However, we observe an increasing (decreasing) on the probability of feeling Spanish (voting regional parties) for individuals with probabilities between 0 and 0.5, and between 0.5 and 1. This is in line with our main findings. Somehow surprising is the fact that, for individuals with a probability of 1 of being sent outside to do their military service, the effect is smaller and in general not statistically different from individuals with a probability of zero. We prefer to remain agnostic about the nature of this finding. On one side, it could be a non-linear effect of the probability of leaving, but we cannot find a reason for it. On the other side, this could be the result of the relatively small number of observations from WSI regions with probability 1 of leaving.

Figure A.6: Estimates allowing non-linear effects



D Heterogeneous Effect by Education

This section looks at the heterogeneous effect by educational level. Specifically, we will look at the effect on individuals with less than tertiary education. These individuals have a higher degree of compliance because they could not postpone their military service for academic duties, which was the most common reason to delay military service.² In addition, the effects of the treatment could be heterogeneous in educational attainment. For example, previous works have found a strong positive correlation observed between regional identity and education (Hierro and Gallego, 2018).

Table A.1 checks the sensitivity of the identity findings by level of education. Column (1) presents our baseline specification to ease the comparison. In Column (2), we restrict the sample to individuals with less than tertiary education. Despite the loss of precision in the estimates after restricting the sample, the point estimates are similar to the estimates in Table 2 in the paper. Specifically, for the individuals from WSI regions, being assigned to another region through the lottery increases Spanish identity by approximately sixteen percentage points. In order to gain precision in the estimates, column (3) presents the results for a model fully flexible in educational level for the complete sample. The effect is significant at 5% significance level (with a point estimate of 0.204) for individuals with lower educational level and with smaller point estimate (0.072) and not statistically significant for those with higher education. The similarity on the estimates of β_1 across specifications, together with a lower expected compliance among individuals with tertiary education, suggests that our baseline estimates are driven by conscripts with lower educational levels.

²Since we do not have information about the time of enlistment, not pursuing post-secondary education reduces the measurement error concerns when using the lottery information for the year of first call. Moreover, although we cannot estimate a first stage, a larger compliance is expected for this sample, that is, a closer effect of the reduced form on the second-stage estimate (according to our calculations, which are above 90%). Finally, even in the hypothetical case of having information about the time of enlistment, this variable might be endogenous since individuals have some power to decide this date by delaying service.

Table A.1: Heterogeneous effect on identity by education level

	(1)	(2)	(3)
	Baseline	Low-ed	All
Out (β_1)	0.166*	0.164	0.204*
	(0.086)	(0.116)	(0.113)
Out x Birth SSI (β_3)	-0.221**	-0.205	-0.240**
	(0.093)	(0.125)	(0.121)
Out x HighEdu (β_4)			-0.132
			(0.245)
Birth SSI x HighEdu (β_5)			0.079
			(0.173)
Out x Birth SSI x HighEdu (β_6)			0.054
			(0.260)
HighEdu (β_7)			-0.064
			(0.166)
Observations	536	392	535
Mean dep. WSI	0.599	0.654	0.603
$\beta_1 + \beta_4$			0.072
$H_0: \beta_1 + \beta_4 = 0$, p-value			0.703

Notes: Each column corresponds to a separate regression. Column 1 is our baseline specification. Column 2 restricts the sample to individuals with low education (without a university degree). Finally column 3 uses the whole sample but adds an interaction between our treatment and education (difference in difference in difference). Birth SSI is a dummy variable equal to one if the individual was born in a region with an SSI. Out is a dummy variable equal to one if the individual was assigned to military service outside his own region. HighEdu is a dummy variable equal to one if the individual has obtained a university degree. All specifications include fixed effects for month of birth and fixed effects for year of birth interacted with region of birth fixed effects. At the bottom of the table we report the sum of coefficients ($\beta_1 + \beta_4$) and the p-value of the test of the sum being equal to zero. Standard errors are clustered at the year of birth-province level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$

E Additional Tables and Figures

Figure A.7: Military areas at the time of the national lottery, for the army (map 1) and the airforce and navy (map 2)



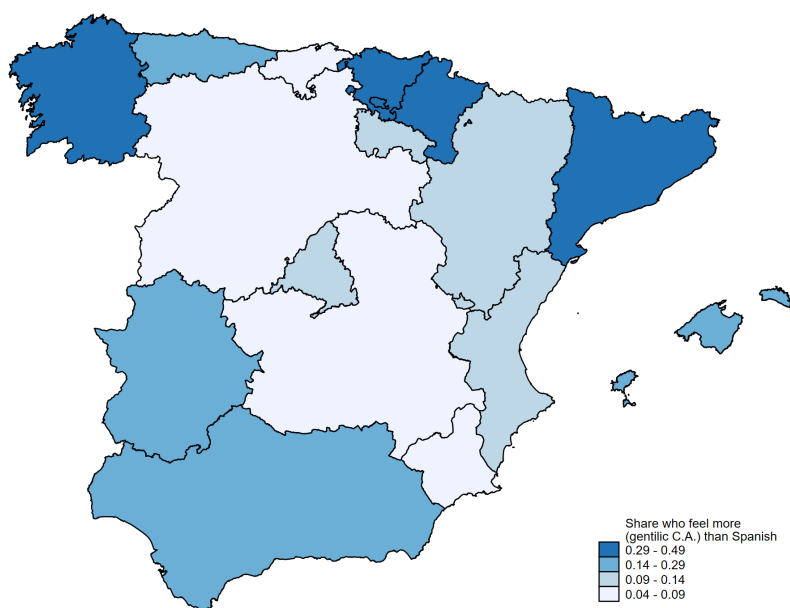
Notes: Military regions as of 1988. Source: La Vanguardia newspaper, November 1988.

Figure A.8: Example of the National Public Lottery

Barcelona			
Demarcación territorial	Fecha de nacimiento	D.N.I.	Total cupo
E.T. Región Centro	Del 26- 4-1962 al 27- 5-1967	36.561.690 46.130.104	3.098
E.T. Región Sur	Del 27- 5-1967 al 31- 5-1965	46.666.545 38.078.879	361
E.T. Región P. Oriental	Del 31- 5-1965 al 25-12-1970	43.407.251 38.111.667	17.970
E.T. Región P. Occidental	Del 25-12-1970 al 1- 1-1970	38.116.783 39.183.364	648
E.T. Zona Baleares	Del 1- 1-1970 al 9- 1-1970	39.345.094 34.737.618	700
E.T. Zona Canarias	Del 9- 1-1970 al 14- 1-1970	34.741.512 39.187.326	431
E.T. Región Sur (Ceuta)	Del 14- 1-1970 al 29- 1-1970	39.346.168 38.112.464	1.242
E.T. Región Sur (Melilla)	Del 29- 1-1970 al 13- 2-1965	38.115.380 46.344.090	1.221
E.A. Primera Zona Aérea	Del 13- 2-1965 al 22- 2-1970	52.191.135 39.346.645	783
E.A. Tercera Zona Aérea	Del 22- 2-1970 al 7- 3-1970	39.348.084 46.574.028	1.154
E.A. Zona Aérea Canarias	Del 7- 3-1970 al 11- 3-1966	46.587.764 39.172.870	286
F.N.-Z.M. Estrecho	Del 11- 3-1966 al 26- 3-1964	43.680.547 35.092.626	1.301
F.N.-Z.M. Mediterráneo	Del 26- 3-1964 al 25- 4-1970	37.183.557 34.737.045	2.638
F.N.-Z.M. Canarias	Del 25- 4-1970 al 25- 4-1970	34.740.467 77.475.614	72

Notes: 1989 replacement lottery (published in November 1988) for the province of Barcelona.

Figure A.9: Share of citizens who feel more gentilic of their region than Spanish



Notes: Our own calculations based on CIS Estudio 2211 - Barometro de Marzo (1996)

Table A.2: Robustness of the national identity results to different definitions of WSI

	(1)	(2)	(3)
	BC+CA+NA+GA	BC+CA+NA	BC+CA
Out (β_1)	0.166*	0.247***	0.236**
	(0.086)	(0.093)	(0.094)
Out x Birth SSI (β_3)	-0.221**	-0.305***	-0.287***
	(0.093)	(0.100)	(0.102)
Observations	536	536	536
Mean dep. WSI	0.599	0.485	0.476
$\beta_1 + \beta_3$	-0.054	-0.058	-0.051
$H_0: \beta_1 + \beta_3 = 0$, p-value	0.130	0.110	0.170

Notes: Each column corresponds to a different definition of WSI. Column 1 is our base specification, where the WSI regions are the Basque Country (BC), Catalonia (CA), Navarre (NA) and Galicia (GA). In column 2, we consider Galicia to be an SSI region, and in column 3, we also add Navarre to the regions with an SSI. Birth SSI is a dummy variable equal to one if the individual was born in a region with an SSI. Out is a dummy variable equal to one if the individual was assigned to military service outside his own region. All specifications include fixed effects for month of birth and fixed effects for year of birth interacted with region of birth fixed effects. Standard errors are clustered at the year of birth-province level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$

Table A.3: Effect on migrating

	(1)	(2)
Out	0.022	-0.002
	(0.034)	(0.051)
Out x Birth SSI		0.033
		(0.062)
Observations	630	630
Mean dep.	0.138	0.138

Notes: The outcome is one if the individual lives in a region different from that of his birth. Birth SSI is a dummy variable equal to one if the individual was born in a region with an SSI. Out is a dummy variable equal to one if the individual was assigned to military service outside his own region. The regression includes fixed effects for month of birth and fixed effects for year of birth interacted with region of birth fixed effects. Standard errors are clustered at the year of birth-province level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$

Table A.4: Robustness to leaving out regions not contained entirely in a military area.

	Spanish identity	
	(1) Whole sample	(2) Robustness sample
Out	0.173* (0.097)	0.174* (0.097)
Out x Birth SSI	-0.229** (0.103)	-0.225** (0.107)
Observations	536	478
Mean dep.	0.834	0.822

Notes: Each column corresponds to a different sample. The first column again shows the results from our main specification. The second column three excludes individuals born in a region that is not entirely contained in a military area. That is, regions that are split in more than one military area. These are the regions of Castilla-La Mancha and Castilla y León. The outcome is an indicator variable that takes the value of one if the individual feels that his Spanish identity is greater than or equal to his regional identity. Birth SSI is a dummy variable equal to one if the individual was born in a region with an SSI. Out is a dummy variable equal to one if the individual was assigned to military service outside his own region. All specifications include fixed effects for month of birth and fixed effects for year of birth interacted with region of birth fixed effects. Standard errors are clustered at the year of birth-province level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$

Table A.5: Classification of parties

Election year	Regionalist 1	Other Parties
2011	CIU, Amaiur, PNV, ERC, BNG, Geroa Bai, CC and FAC.	PP, PSOE, IU (with local alliances), UPyD and EQUO.
2015	ERC, DiL, PNV, EH Bildu and CC.	PP, PSOE, C's, Podemos (with local alliances) and IU.
2016	ERC, CDC, PNV, EH Bildu and CC.	PP, PSOE, Unidos Podemos (with local alliances) and C's.
Election year	Regionalist 2	Other Parties
2011	CIU, Amaiur, PNV, ERC, BNG and Geroa Bai.	PP, PSOE, IU (with local alliances), UPyD, EQUO, CC and FAC..
2015	ERC, DiL, PNV and EH Bildu.	PP, PSOE, C's, Podemos (with local alliances), IU and CC.
2016	ERC, CDC, PNV and EH Bildu.	PP, PSOE, Unidos Podemos (with local alliances), C's and CC.
Election year	Regionalist 3	Other Parties
2011	CIU, Amaiur, PNV, ERC, BNG, Geroa Bai, CC, FAC, ICV and Compromis.	PP, PSOE, IU (without local alliances), UPyD and EQUO (without local alliances).
2015	ERC, DiL, PNV, EH Bildu, CC, En Marea, En Comu, Compromis.	PP, PSOE, C's, Podemos (without local alliances) and IU.
2016	ERC, CDC, PNV, EH Bildu, CC, En Marea, En Comu and Compromis	PP, PSOE, Unidos Podemos (without local alliances) and C's.

Notes: We consider all parties with representation in the Spanish Congress. CIU, ERC, DiL and CDC are Catalan nationalist parties. Amaiur, PNV, Geroa Bai and EH Bildu are Basque nationalist parties, while Geroa Bai operates only in the region of Navarre. CC is a Canarian regionalist party. FAC is an Asturian regionalist party. PP, PSOE, IU, UPyD, EQUO, Podemos and Unidos Podemos are parties present in the entire Spanish territory. ICV is a Catalan party that ran in a coalition with IU. Compromis is a Valencian party that ran in a coalition with EQUO in 2011, with Podemos in 2015 and with Unidos Podemos in 2016. En Comu is the Catalan alliance between IU and Podemos in 2015 and Unidos Podemos in 2016. En Marea is the Galician alliance between IU and Podemos in 2015 and Unidos Podemos in 2016.

Table A.6: Robustness of the voting results to different definitions of regionalist party

	(1)	(2)	(3)
	Regionalist 1	Regionalist 2	Regionalist 3
Out (β_1)	-0.134*	-0.134*	-0.161**
	(0.078)	(0.077)	(0.080)
Out x Birth SSI (β_3)	0.115	0.129*	0.151*
	(0.079)	(0.077)	(0.081)
Observations	536	536	536
Mean dep. WSI	0.204	0.204	0.232
$\beta_1 + \beta_3$	-0.019	-0.004	-0.010
$H_0: \beta_1 + \beta_3 = 0$, p-value	0.251	0.649	0.626

Notes: Each column corresponds to a different definition of regionalist party. Column 1, which is our base specification, uses the definition Regionalist 1. Column two uses Regionalist 2, a more conservative definition of regionalist party, and column 3 uses Regionalist 3, the broadest definition that covers more parties. Null and nonvoters are included as zero in all regionalist definitions. Birth SSI is a dummy variable equal to one if the individual was born in a region with an SSI. Out is a dummy variable equal to one if the individual was assigned to military service outside his own region. All specifications include fixed effects for month of birth and fixed effects for year of birth interacted with region of birth fixed effects. Standard errors are clustered at the year of birth-province level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$

Table A.7: Robustness of identity and voting effects to the 2017 Spanish Constitutional Crisis

	Spanish identity		Voted	
	(1)	(2)	(3)	(4)
	All years	Without 2017	All years	Without 2017
Out (β_1)	0.170*	0.276**	-0.137*	-0.156
	(0.087)	(0.121)	(0.077)	(0.121)
Out x Birth SSI (β_3)	-0.222**	-0.317**	0.116	0.122
	(0.094)	(0.136)	(0.079)	(0.126)
Observations	536	323	536	323
Mean dep. WSI	0.599	0.593	0.204	0.244
$\beta_1 + \beta_3$	-0.052	-0.041	-0.021	-0.034
$H_0: \beta_1 + \beta_3 = 0$, p-value	0.146	0.410	0.215	0.143

Notes: Column 1 repeats the results for our main specification for identity, and column 3 repeats the results of our main specification for having voted in the last Spanish election. Columns 2 and 4 repeat the analysis for each outcome without individuals interviewed in the 2017 survey; this wave was affected by the Spanish Constitutional Crisis following the Declaration of Independence of the Catalan Parliament. Birth SSI is a dummy variable equal to one if the individual was born in a region with an SSI. Out is a dummy variable equal to one if the individual was assigned to military service outside his own region. All specifications include fixed effects for month of birth and fixed effects for year of birth interacted with region of birth fixed effects. Standard errors are clustered at the year of birth-province level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$

Table A.8: SUR models

	Spanish Identity		Vote Regionalist	
	Male	Female	Male	Female
Out (β_1)	0.166** (0.081)	0.018 (0.094)	-0.134* (0.074)	-0.023 (0.068)
Out x Birth SSI (β_3)	-0.221** (0.088)	0.026 (0.100)	0.115 (0.074)	0.006 (0.063)
Observations	1093		1093	
$H_0: \beta_1^{males} = \beta_1^{females}$, p-value	0.257		0.297	

Notes: Columns 1 and 2 show results for a seemingly unrelated regressions specification for Spanish identity using separate equations for males and females. Similarly, columns 3 and 4 repeat the exercise for vote regionalist. Birth SSI is a dummy variable equal to one if the individual was born in a region with an SSI. Out is a dummy variable equal to one if the individual was assigned to military service outside his own region. All specifications include fixed effects for month of birth and fixed effects for year of birth interacted with region of birth fixed effects. At the bottom of the table we report the sum of coefficients ($\beta_1^{male} + \beta_1^{female}$) and the p-value of the test of the coefficients being equal, for each outcome. Standard errors are clustered at the year of birth-province level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$

References

- Guinjoan, Marc and Toni Rodon. 2015. "A scrutiny of the Linz-Moreno question." *Publius: The Journal of Federalism* 46(1):128–142. 3
- Hierro, María José and Aina Gallego. 2018. "Identities in between: Political conflict and ethnonational identities in multicultural states." *Journal of Conflict Resolution* 62(6):1314–1339. 7
- Rodon, Toni and Marc Guinjoan. 2018. "When the context matters: Identity, secession and the spatial dimension in Catalonia." *Political Geography* 63:75–87. 3