Who backs separation? An empirical assessment of the objective and subjective markers of identity in support for secession in Catalonia

Marc Guinjoan

Institutions and Political Economy research Group (IPERG)
Universitat de Barcelona
marc.guinjoan@ub.edu

Abstract: Objective markers of distinctiveness such as language, ethnicity, religion or origin are instrumental in the formation of territorial identities. In turn, territorial identity tends to be the most powerful factor in explaining differences in support for secession. Along with this subjective identity, objective markers of distinctiveness play also a role in explaining support for secession. In this paper I argue that the literature lacks an empirical assessment of how objective and subjective markers of identity are related to support for secession. In particular, I argue that most of the effect of objective markers on support for secession is indirectly expressed via (changes in) identity. Results from survey data from Catalonia (2015-2020, N=30,000), confirm that a large share of the effect of the region's objective markers—language and ancestry—on secession is, as expected, mediated. This evidence has both theoretical and empirical implications for studies on identity formation and support for secession.

Keywords: objective markers, identity, secession, language, ancestry, Catalonia, survey data, mediation.

Social scientists have long discussed how identities are formed. Despite the debate being a lively one and their positions often irreconcilable, most authors agree on the fact that objective markers of distinctiveness play *some* sort of role in the formation of identities. These markers include language, religion, ethnicity and origin, among others. In parallel, a growing branch of the political science literature has focused its attention on identifying the individual-level mechanisms that might explain why some individuals favour their region becoming a new independent state while others do not. Despite the explicative factors varying considerably across cases, one's attachment to a nation or to a region—i.e. an individual's identity—has been recognised as a major force. Hence, individuals who feel closer to a region or identify themselves with regional values are more likely to favour secession than individuals who identify with the nation-state.

While the role of objective markers of distinctiveness on the formation of a national-regional identity is clear, as is the impact of identity on support for secession, the literature still lacks an empirical assessment regarding the extent to which objective markers of distinctiveness have a direct effect on support for secession. The conundrum is the following: do cultural and ethnic attributes directly impact an individual's territorial preferences? Or, rather, is it only through the formation of an identity that these markers have an impact on support for secession? By addressing this question, in this paper I aim to offer a bridge between the literature on identity formation and that on support for secession, and to present an original, empirical assessment of the determinants of support for secession that incorporates existing theories on identity formation. My argument is that, the impact of the objective markers of distinctiveness

on support for secession will be, rather than a direct effect, a mediated one. Schematically:

Objective markers → National-regional identity → Support for secession

To address the extent to which objective markers *directly* explain support for secession, I employ Catalonia as an ideal case study. Using data from the Catalan polling institute from 2015-2020 (N=30,000) and a series of Structural Equation Models, I will be able to determine the percentage of the effect on support for secession of objective regional markers—in the case of Catalonia, regional language and ancestry—that is expressed through national-regional identity.

The results of my analyses show that more than 90 percent of the change in support for secession in Catalonia due to having Catalan as one's first language or having Catalan ancestries is expressed via the configuration of one's national-regional identity. Overall, this evidence reveals the complexity of measuring and conceptualising national-regional identity, and has both theoretical and empirical implications for studies that deal with identities, ethnicities and, ultimately, support for secession.

IDENTITY FORMATION AND SUPPORT FOR SECESSION

How is identity formed?

The formation of nations is not denoted by concrete historical milestones, but is instead a complex process that includes several concurring, non-deterministic, contextual and dynamic factors. Even though it is not the purpose of this research to embark on a thorough revision of the literature on the formation of identities—and, even less, to take

sides in the ongoing ontological debates on the subject—I subscribe to the idea that, in the formation of an identity, both *objective* ethnic markers and *subjective*, strategic and opportunistic issues (in the hands of political entrepreneurs) have *some* impact. Several social scientists have seconded this idea.

For Hroch (1985), the configuration of a national identity begins when a collective raises awareness of its 'common traits', including the memory of some common past, the existence of linguistic or cultural ties that enable a higher degree of social communication within the group, and a conception of equality between the members of the group. Similarly, for Hobsbawm (1990), for a national idea to be formed—a 'popular proto-nationalism'—a collective requires a language, an ethnicity, and a religion—along with the awareness of belonging, or having belonged to, a lasting political entity. This idea is ingeniously summarised by Hall (1997, 36) in his assertion that "enunciation comes from somewhere."

However, when considering the formation of a subjective identity, the *objective* conditions of existence are, for most scholars, not enough. Indeed, discursive approaches see identification as a non-determined and conditional construction which is lodged in contingency (Hall 2011), and thus understand the formation of an identity as a dynamic process related both to the existence of cultural meanings, but also to wider historical, social and material circumstances (Verkuyten 2005; Hall 1997). Although identities often invoke a common past to which the community still corresponds, they are "about using the resources of history, language and culture in the process of becoming rather than being" (Hall 2011, 4). Unsurprisingly, the boundaries between objective and subjective conditions are not easy to draw, and the various ontological

approaches recognise the existence of objectivity in different ways and attribute diverse meanings to it—from recognising objective realism, to the credence that the physical world is itself only a social construction (c.f. della Porta and Keating 2008).

Several authors from the world of political science have addressed the relationship between objective and subjective markers too. To mention a few, Margalit and Raz (1990) assert that it is the presence of a common character—including features such as a national cuisine, architectural style, a common language, literary or artistic tradition, similar customs, dress, etc.—that determines a nation's unique identity. In addition, Sorens (2005) claims that the presence of a regional language—together with a previous history of independence—is the activator of ethnic identity (see also Moore 2001; or Griffiths 2016). All in all, without denying the fact that a subjective identity is, to some extent, a social construct—which, in turn, may also affect the determination of what is perceived as *objective*—in this research I will assume that objective identities are as much related to the tradition itself as to the invention of a tradition (Anderson 1983; Gilroy 1994) and, as such, objective markers of identity—at least in the case being studied and of the objective traits under consideration—can, to a certain degree, be objectifiable.

Which markers of identity should be considered?

Objective markers of distinctiveness vary noticeably depending on the case being studied (see Griffiths and Martinez 2020). Hence, for instance, in Northern Ireland (United Kingdom), in Tibet or Xinjiang (China), in Kachin (Myanmar) or in Chechnya (Russia), religion is the main catalyser of regional identity. In Ahwaz (Iran)—along with religion—, in Western Sahara (Morocco) or in Kurdistan (Iran, Turkey, Iraq, and

Syria), ethnicity is the foremost marker of identity. Language is also a crucial marker of identity in Kurdistan, in Crimea, Donetsk and Luhansk (Ukraine), but also in Quebec (Canada), Flanders and Wallonia (Belgium), Aaland (Finland) South Tyrol (Italy) or in Catalonia and the Basque country (Spain). Finally, ancestry is in certain cases an encompassing factor that may include various markers of identity. In other words, the regional lineage will often be closely associated with following the local religion, speaking the regional language or belonging to the region's ethnic group.

Despite wide variations in the determinants of a regional identity, in this study I focus on *language* and *ancestry*, for several reasons. First, most anthropologists have relied on these two variables for determining whether individuals should be classified as 'alike' or not (Labov 1973). Second, in the case being studied, Catalonia, language and ancestry are the most relevant markers of distinctiveness. Third, individual-level data containing regional identification variables is extremely rare—particularly in non-Western countries—and the focus of a study is largely constrained by the availability of data.

The relationship between *language* and identity has received considerable attention in the anthropology, sociolinguistics and humanities literature. In Kamusella's words (2001, 235), "language is the very 'stuff' of culture as well as the instrument of communicating and reaffirming cultural difference *vis-à-vis* other cultures". There are at least four different mechanisms that link language and identity. First, among the many symbolic resources available for the promotion of identity, language is the most flexible and pervasive. Language not only allows individuals to perceive themselves as part of a group, but also produces social distance between those who perceive themselves as

different (Bucholtz and Hall 2004, 367). Second, language does not only entail a functional practice, but also becomes another form of social activity, and shapes an individual's *habitus* (Bourdieu 1977). Third, language shapes different frameworks of reference, driving citizens' media and cultural consumption patterns, and triggering different identity-contexts (Barceló 2014), which may eventually modify an individual's identity (Adoni, Cohen, and Caspi 2002). Last, languages have become instruments in the pursuit of state nationalism (c.f. Iwamoto 2005), as well as instruments of differentiation in the hands of local and regional elites (see, for instance Breton 1978 for the case of Quebec; Jaffe 1999 for Corsica; or Hannan 1996 for Silesia). Hence, the politicisation of the language issue in minority nationalist regions may serve the purpose of establishing a dichotomy between social identities by reducing complex social variability (Bucholtz and Hall 2004), and may also lead to changing stances on national-regional identification.

The relationship between an individual's *ancestry* and her identity is also widely reported in the literature. Ancestry is the 'inherited' element of ethnicity (Brubaker et al. 2007), the belief in common descent or kinship (Jiménez 2010). Research on immigration and ethnic groups has shown that in almost all countries, immigrants and cultural minorities have lower host national identification than majority group members (e.g., Phinney et al. 2006; Staerkle, Sidanius, Green, and Molina 2010). Several mechanisms explain why people born outside the nation (region), or whose parents were born outside it, feel that they belong less to the nation (region). First, foreigners—or their sons—may deliberately choose to exclude themselves from the identity associated with the majority because they have less (or even a complete lack of) motivation to claim to be members of a group to which they feel they do not belong (Bond 2006).

Second, claims about identity are delimited by the actions and reactions of others (Bechhofer et al. 1999, 527); hence, any feelings of identification held by those who are different—in terms of ancestry, in this case—may be problematized by the beliefs of the majority about the validity of these claims (Bond 2006).

*What explains support for secession?*

There is widespread agreement in the literature that, in minority nationalist regions where different national and regional identities coexist, the more an individual feels attached to the region rather than to the nation, the more likely it is that she develops a preference for higher degrees of decentralization or secession (Lluch 2014).[1] Previous works have found a strong link between regional identity and secessionist attitudes in Catalonia (Muñoz and Tormos 2015; Guinjoan and Rodon 2016a; Rodon and Guinjoan 2018; Serrano 2013), in the Basque country (Costa-Font and Tremosa-Balcells 2008; Serrano 2019), in Scotland (Muro and Vlaskamp 2016; Liñeira and Cetrà 2015), in Quebec (Nadeau and Blais 1991), in Flanders (Cetrà 2019), in (Turkish) Kurdistan (Sarigil and Karakoc 2016) or in Indonesia (Webster 2007), among others. For instance, in Catalonia, Serrano (2013) has shown that individuals that feel 'only Catalan' are over four times more likely to support independence than those who feel 'only Spanish'.

In addition to national-regional identity, correlational evidence has shown that language and ancestry are also related to support for secession. In Catalonia, for instance, where Catalan is the mother tongue of almost half the population, 71 percent of those who speak Catalan as their first language support secession, while the percentage drops to 28 percent for bilingual individuals, and plummets to 11 percent for Spanish speakers (Guinjoan and Rodon 2016a). In Quebec, the percentage of French speakers who

favoured the region seceding in 2015 was around 35 percent, while among English speakers this was just 4 percent (Blais 2010). Similarly, *ancestry* registers as a strong correlate of support for secession. Costa-Font and Tremosa-Balcells (2008, 2470) sustain, for the Basque country, that an individual's birthplace is "seen as a proxy of individuals' benefits from secession as secession might arguably lead to a transfer of central power to the Basque Country". In Catalonia, in 2014, individuals born in the region whose parents were both also born there had a 71 percent likelihood of favouring secession. This was in contrast to people born in Catalonia with one or neither parent born in the region, whose chances of supporting secession dropped to 49 and 20 percent, respectively. Finally, the likelihood of those born outside Catalonia supporting succession was only 18 percent (Guinjoan and Rodon 2016a).

Even though correlational evidence relates objective markers of identity with support for secession, I sustain that there are no theoretical arguments to expect such distinctive markers to explain support for secession, other than through the configuration of a regional identity. The mere presence of regional markers will not translate into preferences for secession if this is not channelled through an acquired (regional) identity. Hence, whether an individual born in a nationalist region and who speaks the local language will eventually support the territory's secession will depend, to a large extent, on the acquisition of a regional identity. If she develops a dual or predominantly national identity, it can hardly be expected that language and ancestry alone will explain preferences for secession. Figure 1 summarises my theoretical expectations.

Figure 1. The mediated relationship between language and ancestry on support for secession
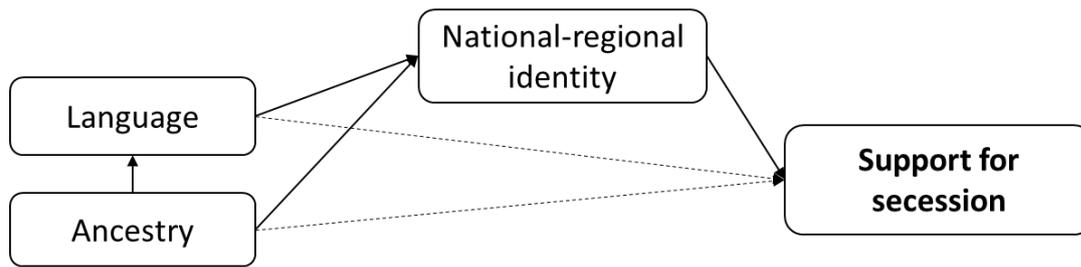
The solid line linking language and ancestry with national-regional identity, and linking the latter variable with support for secession portrays the *indirect effect* of language and ancestry on preferences for secession. The indirect effect of language (ancestry) on support for secession must be understood as the increase we would see in the support for secession while holding language (ancestry) constant and increasing identity to whatever value this would reach under a unit increase in language (ancestry) (Pearl and Mackenzie 2018, 298). The dotted line linking language and ancestry with support for secession displays the *direct effect* of the first two factors on the third. This is the part of the effect of language (ancestry) which is not mediated by national-regional identity, and I expect its impact on support for secession to be close to zero.

DATA AND METHODS

The study is focused on Catalonia between 2015 and 2020. Catalonia—and this specific period of time—provides an ideal setting to study how an individual's ancestry and her language affects their national-regional identity and their support for secession. There are several reasons for this. First, Catalonia is a minority nationalist region (Keating 1996) and, as such, Catalan and Spanish identities are conceived as having some degree of exclusivity for a relevant part of the population (Guinjoan and Bermúdez 2020). Second, unlike other regions that hold nationalist claims, the percentage of people born

outside Catalonia is particularly high, thus granting important variation. According to data from the Catalan Statistics Institute (IDESCAT n.d.), in 2019, 16.6 per cent of the population resident in Catalonia was born in other parts of Spain and an additional 19.1 per cent was born outside Spain. The distribution of spoken languages in Catalonia also provides us with key variation: in 2013, Catalan was the language of identification for 36.4 per cent of the population, while for 47.6 per cent, it was Spanish. Individuals who identify with both languages account for 7 per cent of the population.

Third, support for secession and national-regional identity during the 2015-2020 period in Catalonia have suffered relatively small oscillations. This contrasts with the previous decade, where both support for secession and the percentage of people who identified as exclusively Catalan skyrocketed. In fact, data from the CEO reveals that the standard deviation in support for secession is 2.2 between 2015 and 2020 period (min=41.1, max=49), but 10.5 between 2006—when the CEO was founded—and 2013 (for a 4 categories question, min=14, max=48.5). Similarly, the standard deviation in Catalan exclusive identities is 2.2 for the later and 4.7 for the earlier period, (min=22.3, max=28.6; and min=14.8, max=31.3, respectively). Thus, by restricting the analysis to this period I can reduce potential sources of statistical disturbance and assure that the analysis is done during a period with relatively stable political opinions.

Fourth, the debate over Catalan secession has been prominent during this period (Cuadras-Morató 2016), thus making it a hot issue for public debate. In fact, since late 2014 and until 2018 Catalonia has experienced a set of polarising events that have often brough the conflict to the international scene. In November 2014 the Catalan government organised a popular consultation on independence which was deemed

illegal by the Spanish Courts (c.f. Guinjoan and Rodon 2016a). The impossibility to hold a legal referendum encouraged the Catalan government to call for early regional election in September 2015, which was announced as a plebiscite for independence. Despite pro-secession parties gained a majority of the seats in the Parliament, they fell short of achieving a majority of the votes (47.7 percent). Under these circumstances, the Catalan government agreed to hold a new and definitive referendum on secession, irrespective of the central government's endorsement. The referendum was announced on Sunday October 1, 2017, and the Spanish Constitutional Court soon declared it unconstitutional. During September 2017 the Spanish government deployed more than 10,000 police officers in Catalonia. On October 1, Spanish police squads were sent to several locations to stop the referendum. Finally, in many polling stations, the police attempts to halt the vote exploded into violence. More than a thousand people were injured (c.f. Barceló 2018; Balcells, Dorsey, and Tellez 2020; Rodon and Guinjoan 2021). These events finally resulted in the suspension of the Catalan autonomy and the call for early elections in December 2017—that, again, were won by pro-secession parties—as well as the imprisonment of most regional ministries from the Catalan government on October 1, and some civil society leaders. Overall, the saliency of the debate during this period is important as this reduces uncertainty about the issue and decrease weak or elusive preferences towards independence (Behrens et al. 2007).

The dependent variable is the individual's stance on the Catalan secession issue. The question asks whether she wants Catalonia to become a State. Respondents were consistently given the following options: "Yes", "No", as well as the "Don't know" and "No answer" categories (not read out). The outcome of interest is a dichotomous variable, with 1 being in favour of Catalonia becoming a State and 0 against it.

The mediator variable is an individual's national-regional identity. I rely on one of the most common measure of identity in empirical analyses: the Linz-Moreno question (Moreno 1995; Guinjoan and Rodon 2016b). This measure asks respondents to choose a single identity from a range of options along a unique dimension whose extreme points are "I feel only [national]" (value 1) and "I feel only [regional]" (value 4). Previous evidence from Catalonia (Serrano 2013; Rodon and Guinjoan 2018; Muñoz and Tormos 2015) shows that individuals who feel "Only Spanish" and "More Spanish than Catalan" have the same positioning towards the Catalan secession debate and, consequently, I have collapsed these two categories into a single group.

The two main independent variables are an individual's language and her Catalan ancestry. Language is a dichotomous variable that identifies which language the respondent first spoke at home when she was a child, where 1 corresponds to Catalan, and 0 to other languages or combination of languages. By using a variable that identifies the first language an individual spoke at home instead of her current language of identification, I am preventing the appearance of reverse causality problems due to the language of *identification* being correlated with an individual's territorial preferences. Ancestry is a variable identifying whether the parents of the individual were both born in Catalonia (1), if just one of them was born in Catalonia (0.5) or none of them (0).[2] The empirical model also includes controls for female; age (continuous variable); ideology (from 0 to 10) and education (1 primary studies, 2 secondary studies, 3 university studies). Finally, to account for the heterogeneous effects of the local environment on support for secession (Rodon and Guinjoan 2018), I include a control for the identity context in which an individual often interacts. This variable—extracted

from the 2011 Census—is coded at the contextual level and measures the percentage of individuals in each of the 42 counties (*comarca*) in Catalonia that are able to speak Catalan. Finally, since data on the percentage of people who speak Catalan is available for each municipality in Catalonia, this has been matched at this level of aggregation for those 23 cities in Catalonia for which the CEO registers the exact municipality where the survey took place; in the remaining cases, data has been matched at the county level. Appendix A1 in the Supplemental Appendix provides information on the specific wording of each of the questions included in the study, and Appendix A2 summarises the statistical descriptives in the empirical analysis.[3]

To assess the impact of language and ancestry on support for secession via national-regional identity, I rely on a series of Structural Equation Models (SEM). SEM models offer more reliability and flexibility for testing posited arguments than regression or mediation effect models. In this case, a SEM model serves for three different purposes. First, it allows estimating the effect of language and ancestry on support for secession contemporaneously, mediated by national-regional identity. Second, it allows the estimation of the effect of language and ancestry on the various categories of national-regional identity—by using a multinomial logistic model—, and allows calculating the percentage of the effect that is mediated by each identity category. Finally, it allows the creation of a causal arrow linking an individual's ancestry and her language, thus controlling for a crucial precedent of language.[4] The estimation of the model is performed by Maximum Likelihood, using a multinomial logistic regression for the mediator (national-regional identity) (step 1)—a Brant Test (Brant 1990) shows that the parallel regression assumption has been violated ($Chi^2= 30.51$ for language, 13.10 for the ancestry: 81.21 overall), thus making it inappropriate to use an ordinal logistic

regression model. In step 2, I use a logistic regression for the dependent variable (support for secession). Finally, the data is clustered by survey and standard errors are bootstrapped in order to properly estimate direct and indirect effects.

The data come from the regular barometers put together by the Catalan government's polling institute, the Centre d'Estudis d'Opinió (CEO n.d.). In particular, I employ a pool of eighteen different surveys and almost 30,000 different individuals interviewed from February 2015 to February 2020 using a face-to-face interview method and stratification by province and size of municipality.[5] Quotas were introduced for sex, age and origin to match the Catalan population.[6]

EMPIRICAL EVIDENCE

Table 1 presents the results of a series of SEM models and Figure 2 plots the predicted values for the identity placement (step 1) and the predicted support for secession (step 2) against the different levels of language and ancestry. Finally, Figure A3a in the Supplemental Appendix A3 plots the causal diagram with the estimated coefficients for each path. Models 1a to 1c in Table 1 assess the impact of language and ancestry on the different categories of national-regional identity (Only Spanish and more Spanish than Catalan reference category) using a multinomial logistic regression model. The two variables are positively correlated with national-regional identity, and the coefficients increase as we move away from Spanish identities. The top left panel in Figure 2 shows the likelihood of being placed in an identity group by language. Moving from an individual who has Spanish as her first language to one whose first language is Catalan increases, *ceteris paribus*, the likelihood of feeling predominantly Catalan from 0.226 to 0.319, and for only Catalan individuals, from 0.153 to 0.443. In contrast, changing from

Spanish to Catalan decreases the probability of having an exclusive Spanish identity from 0.131 to 0.028, and, for dual-identity individuals, from 0.490 to 0.210.

Ancestry is also a crucial determinant of national-regional identity, yet its impact is lesser in size than language. As shown in the top right panel in Figure 2, having the two parents born in Catalonia, when compared to someone whose parents were born outside the region, increases the likelihood of feeling more Catalan than Spanish from 0.235 to 0.292, and from 0.179 to 0.377 for only Catalan individuals. The slope is negative for predominantly Spanish individuals (from 0.142 to 0.035) and for dual-identity individuals (from 0.444 to 0.296).

In Model 2, the dependent variable is support for secession. The main explicative variable is national-regional identity. While the likelihood of being in favour of secession among predominantly Spanish individuals is as low as 0.099, this probability reaches 0.951 among exclusively Catalan individuals (0.183 for dual-identity individuals and 0.736 for predominantly Catalan individuals). Language and ancestry are also statistically significant variables, whose impact is, in this case, much smaller.

The bottom panels in Figure 2 show the likelihood of being in favour of secession by identity group and language (left panel) and ancestry (right panel). As stated earlier, the gap displayed by language (ancestry) in each identity group represents the direct effect of this variable on support for secession, i.e. the unmediated effect via the national-regional identity. A larger gap between the two slopes (in the same identity category) entails a larger direct impact of language (ancestry) on support for secession. The bottom left panel in Figure 1 shows that the difference by language in support for

secession among predominantly Spanish and among exclusively Catalan individuals is around 2-3 percentage points (in favour of those who have Catalan as a first language). Among dual-identity and predominantly Catalan individuals that language gap widens to 6.0 and 7.5 percent points, respectively, thus showing a stronger direct effect of language on support for secession among these intermediate groups.

The bottom right panel displays the gap in support for secession by ancestry. The observed changes are almost identical to the ones in language. For predominantly Spanish and exclusively Catalan individuals the gap between those whose two parents were born in and outside Catalonia is also around 2-3 percentage points, while among dual-identity and predominantly Catalan individuals it reaches 6.0 and 7.8 percentage points, respectively. Finally, Model 3 in Table 1 shows that ancestry is, as expected, a strong predictor of language.

This evidence shows that, while language and ancestry are powerful predictors of an individual's national-regional identity, variation in support for secession is mostly accounted for by national-regional identity, leaving a minor role to language and ancestry.
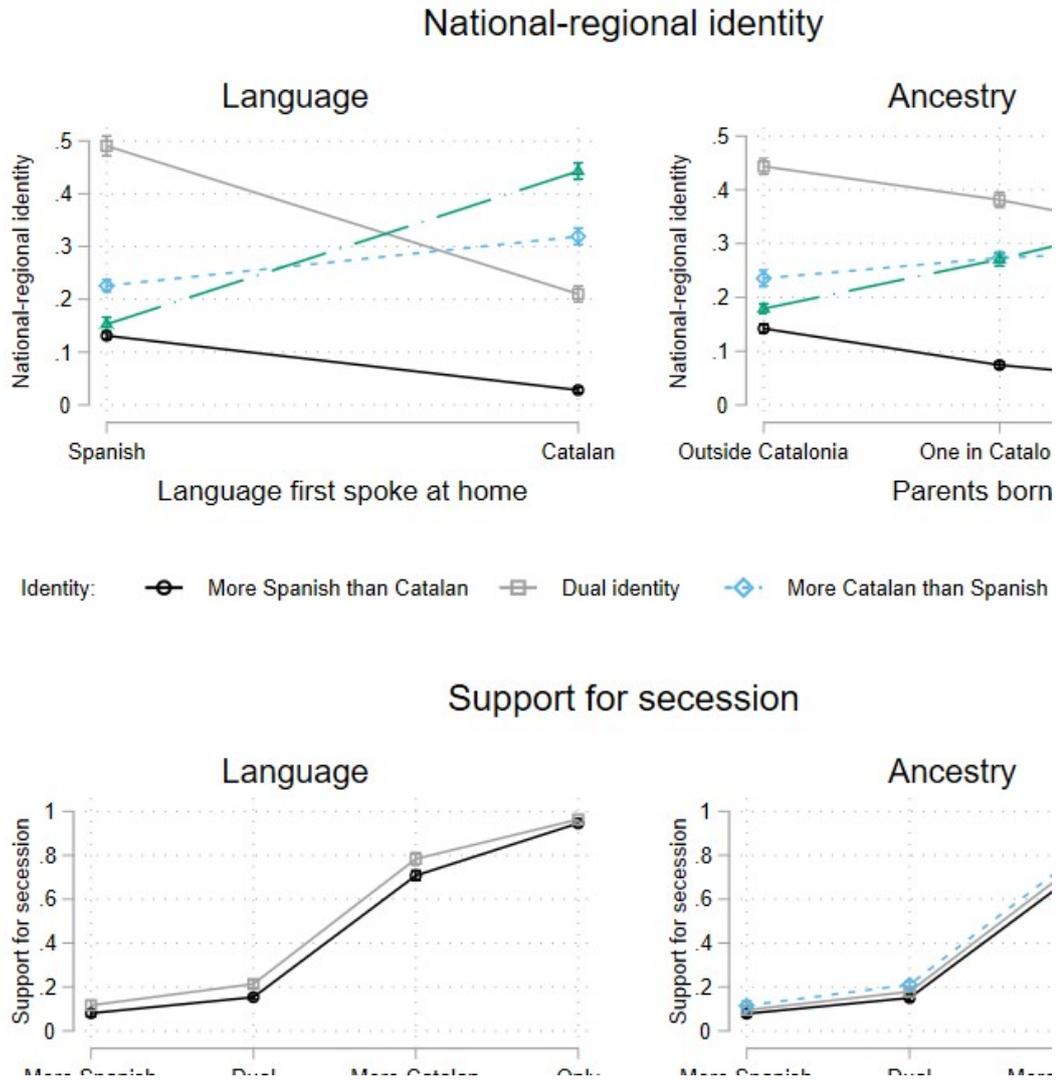
Table 1. Structural Equation Model

	(1a) Identity: Dual	(1b) Identity: More Catalan	(1c) Identity: Only Catalan	(2) Secession	(3) Language
Language:	0.789***	2.076***	2.871***	0.425***	
	(0.102)	(0.096)	(0.084)	(0.070)	
Ancestry	1.068***	1.857***	2.502***	0.435***	4.754***
	(0.102)	(0.088)	(0.085)	(0.062)	(0.045)
Identity					
Only Spanish /More Spanish				[Ref.]	
Dual identity				0.751***	
				(0.108)	
More Catalan				3.511***	
				(0.128)	
Only Catalan				5.555***	
				(0.158)	
Context				2.726***	
				(0.335)	
Ideology				-1.764***	
				(0.148)	
Education					
Primary				[Ref.]	
Secondary				0.023	
				(0.046)	
University				0.103	
				(0.068)	
Female				-0.021	
				(0.043)	
Age				-0.011***	
				(0.001)	
Constant	0.982***	-0.182***	-0.980***	-3.589***	
	(0.026)	(0.042)	(0.053)	(0.269)	
Cluster (survey)			Yes		
Observations			25,777		
Log lik.			-53967.816		

Note: Reference category for the identity variable: Only Spanish/More Spanish than Catalan
Standard errors in parentheses
* $p < .05$, ** $p < .01$, *** $p < .001$

Figure 2. Identity placement and support for secession

National-regional identity

Language

Ancestry

Identity: —⊖— More Spanish than Catalan   —⊟— Dual identity   ·◇· More Catalan than Spanish

Support for secession

Language

Ancestry

Note: Whiskers display 95 percent CI

Finally, I compute the direct effect of language and ancestry on support for secession, the indirect effect via national-regional identity, the total effects (the sum of the direct and indirect effects), and the percentage of the total effect that is mediated via national-regional identity. Recall, once again, that the indirect effect has to be understood as the increase we would see in the support for secession while holding the language (or ancestry) constant and increasing identity, while the percentage of the mediated effect is the ratio between this change and the total change—note that, despite being unusual, the

percentage of the mediated effect can take a value of above 100 percent if the indirect and the direct effects are linked with the outcome in the opposite direction. Only Spanish and more Spanish than Catalan individuals are, again, the reference category for the identity variable and, hence, the estimation of the indirect and total effects has to be understood as a change from this reference group towards each of the other identity groups.

Table 2 largely confirms the evidence found in the previous empirical assessment. A large percentage of the effect of both language and ancestry are mediated through national-regional identity. Remarkably, the percentage of the effect of language and ancestry that is expressed via identity increases as we move towards more Catalan identities. Hence, in looking at the change in support for secession when moving from an individual who does not speak Catalan as her first language to someone who does, we see that 58 percent of the change can be accounted for by an increase in identity from someone who feels predominantly Spanish to someone who holds a dual identity. This mediated percentage goes up to 95 percent and 97 percent for changes in identity from someone who feels predominantly Spanish to predominantly Catalan, and from predominantly Spanish to exclusively Catalan, respectively.

When explaining changes in support for secession when moving from someone whose parents were born outside Catalonia to someone whose parents were born in the region, 65 percent of the change is explained by a move from holding a predominantly Spanish to a dual-identity, and 94 percent and 97 percent by a change from a predominantly Spanish to a predominantly Catalan identity, and to exclusively Catalan, respectively. This evidence is consistent with Rodon and Guinjoan (2018), who show that dual-

identity individuals in Catalonia have less deterministic stances on the territorial debate

and are thus more elastic towards the presence of external factors.

Table 2. Direct, indirect, and total effects for the fitted model for the language and the ancestry

Effect	Language		Ancestry	
	Observed Coef.	Std. Dev	Observed Coef.	Std. Dev
Direct	0.425*	0.070	0.435*	0.063
Indirect				
Only Sp./More Sp.	[Ref.]		[Ref.]	
Dual identity	0.593*	0.108	0.803*	0.146
More Catalan	7.291*	0.411	6.521*	0.401
Only Catalan	15.946*	0.480	13.900*	0.659
Total effect				
Only Sp./More Sp.	[Ref.]		[Ref.]	
Dual identity	1.018*	0.150	1.237*	0.166
More Catalan	7.716*	0.443	6.956*	0.429
Only Catalan	16.371*	0.499	14.334*	0.665
Percent indirect				
Only Sp./More Sp.	[Ref.]		[Ref.]	
Dual identity	58.2%*	0.047	64.8%*	0.050
More Catalan	94.5%*	0.008	93.8%*	0.008
Only Catalan	97.4%*	0.004	97.0%*	0.004

N=29,771
* $p < .001$

Finally, Appendix A4 in the Supplemental Appendix runs a series of robustness checks

that largely confirm the mediating role of national-regional identity on the relationship

between language/ancestry and identity, and support for secession. First, results remain

unaltered when I consider two potential confounders of identity and secession in

Catalonia such as the (linguistic) context where an individual lives (Barceló 2014; Rico

and Jennings 2012; Rodon and Guinjoan 2018) and an individual's ideology (Galais and

Serrano 2020; Liñeira, Muñoz, and Rico 2020). Second, the disaggregation of the

empirical analysis by years displays low variation in the percentage of the mediated

effect by language and ancestry across years and, what is more, this does not seem to follow any specific pattern—thus providing enough evidence that results are not contingent on the particular political events that took place during the period analysed. Third, the use of an ordinal measure of national-regional identity largely confirms the existence of a strong mediated effect via changes in national-regional identity of both language (90 per cent of the total effect) and ancestry (88 per cent) on support for secession. Fourth, I have also replicate previous analyses by using a 4-item variable of language (Catalan, Spanish, bilingual Catalan-Spanish, and other languages or combinations). The results, again, support the idea that most of the impact of each language category in support for secession is accounted for through changes in national-regional identity. Fifth, I include a causal arrow between support for secession and national-regional identity that accounts for the non-recursive relationship between identity and territorial preferences (Tormos, Muñoz, and Hierro 2015). Despite the methodological difficulties in estimating such a model in observational studies, the evidence confirms the existence of a strong mediated effect of both language and ancestry on support for secession via national-regional identity—even if the mediated effect is almost halved. Last, I rely on an alternative estimation procedure for computing direct and indirect effects when there are variables to decompose that are categorical (Buis 2010; Erikson 2015). Again, the results, displayed in Appendix A4 are analogous to all previous evidence.

DISCUSSION

National-regional identity, a *subjective* measure that identifies the degree to which individuals feel attached to the region and to the nation they live in, has largely been considered the most powerful factor for explaining individual-level differences in

support for a region's secession. Along with identity, *objective* regional markers (be they language, ancestry, religion, ethnicity …) have also been shown to be important predictors of both the national-regional identity, as well as the territorial preferences. In this paper I have empirically proved that, aligned with most social scientists' theories, most of the effect of the objective markers on support for secession is expressed through (changes in) national-regional identity, rather than directly.

Using data from Catalonia (2015-2020)—where the most prominent regional markers are being a native speaker of Catalan and having both parents born in the region—I have shown that more than 90 percent of the effect of both language and ancestry on support for secession is expressed through the configuration of national-regional identity—at least for predominantly and exclusive Catalan identity groups. Several empirical checks, using different methodological strategies and subsamples of data, have confirmed the robustness of my results.

Beyond theoretical and conceptual considerations, the existence of a temporal sequence between objective and subjective markers of identity has important empirical consequences (c.f. Pearl and Mackenzie 2018; Keele, Stevenson, and Elwert 2020). As antecedents of national-regional identity, the evidence presented in this study leans in favour of the inclusion of language and ancestry in models that assess the impact of identity on support for secession. Failing to do so will lead researchers not only to an inaccurate conceptual understanding of the *true* relationship between identity and support for secession, but also to biased estimates. By contrast, when the object of interest is language or ancestry (or any other objective characteristic), the inclusion of national-regional identity would trump a proper assessment of the effect of regional

markers on support for secession. Indeed, a very common approach in social sciences is to condition on post-treatment variables—i.e. variables included in the causal path between the independent variable of interest and the outcome variable. Then, fixing some consequences (identity) of the treatment (language, ancestry) will result in seriously biased estimates of the direct effects of interest (Acharya, Blackwell, and Sen 2016; Montgomery, Nyhan, and Torres 2018).

Even though the evidence is consistent with previous theoretical arguments and my own expectations, several issues may interest future scholars. First, using Catalonia as a case study is not devoid of important implications. In this region, having Catalan as the first language is the most important predictor of national-regional identity. In other regions, however, the role of the regional language is undoubtedly different. For instance, support for secession is much higher in Scotland than in Quebec, even though Gaelic and French are the mother language of 1 and 90 percent of the respective populations. Second, further research should not only test the external validity of the posited arguments and findings in other regional settings where the distribution of languages and ancestry is different from the Catalan case, but should also focus on other markers of identity such as religion or ethnicity. By doing this, the analysis could be expanded to other non-western countries, where a similar mediated relationship can be expected, yet potentially with dissimilar markers of identity. Third, in methodological terms, my empirical results are estimated using observational data. By trying to adjust for observed coefficients I have aimed at making the treatment states as-if random, yet the true causal effect would be better identified with a fully randomised manipulation of the treatment variables. Experimental techniques would also be suited to disentangle an under-researched topic in the literature such as the non-recursive relationship between

national-regional identity and support for secession (Tormos, Muñoz, and Hierro 2015). Similarly, my empirical estimations are based on a theoretical model that assumes that the posited argument represents the true causal structure. Such structure may, however, be refined or even called it into question, as the empirical results might then be too. This becomes particularly dramatic if we bear in mind the complex relationship between objective and subjective markers of identity—and not to mention the ontological disputes around the existence and meaning of 'objectivity.'

Overall, this research represents a first step in empirically reporting and quantifying the existence of a mediated relationship between objective markers, subjective and strategic issues, and support for secession. The evidence found invites us to bridge the literatures on identity formation and support for secession and allows us to better understand the theoretical and empirical underpinnings of support for secession and their causal structure.

REFERENCES

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. Explaining Causal Findings without Bias: Detecting and Assessing Direct Effects. *American Political Science Review* 110 (3): 512–29. doi:10.1017/S0003055416000216.

Adoni, Hanna, Akiba A. Cohen, and Dan Caspi. 2002. The Consumer's Choice: Language, Media Consumption and Hybrid Identities of Minorities. *Communications* 27 (4): 1613–4087. doi:10.1515/comm.2002.001.

Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. New York: Verso.

Balcells, Laia, Spencer Dorsey, and Juan F. Tellez. 2020. Repression and Dissent in Contemporary Catalonia. *British Journal of Political Science*, 1–9. doi:10.1017/S0007123420000307.

Barceló, Joan. 2014. Contextual Effects on Subjective National Identity. *Nations and Nationalism* 20 (4): 701–20. doi:https://doi.org/10.1111/nana.12080.

———. 2018. Batons and Ballots: The Effectiveness of State Violence in Fighting against Catalan Separatism. *Research & Politics* 5 (2). doi:10.1177/2053168018781747.

Behrens, Timothy E J, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth. 2007. Learning the Value of Information in an Uncertain World. *Nature Neuroscience* 10 (9): 1214–21. doi:10.1038/nn1954.

Blais, André. 2010. Making Electoral Democracy Work. *Electoral Studies* 29 (1): 169–70. doi:10.1016/j.electstud.2009.05.006.

Bond, Ross. 2006. Belonging and Becoming: National Identity and Exclusion. *Sociology* 40 (4): 609–26. doi:10.1177/0038038506065149.

Bourdieu, Pierre. 1977. The Economics of Linguistic Exchanges. *Social Science*

Information 16 (6): 645–68. doi:10.1177/053901847701600601.

Brant, Rollin. 1990. Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics* 46: 1171–1178. doi:10.2307/2532457.

Breton, Albert. 1978. Nationalism and Language Policies. *The Canadian Journal of Economics / Revue Canadienne d'Economique* 11 (4): 656–68. doi:10.2307/134371.

Brubaker, Rogers, Margit Feischmidt, Jon Fox, and Liana Grancea. 2007. *Nationalist Politics and Everyday Ethnicity in a Transylvanian Town*. Princeton, NJ: Princeton University Press.

Bucholtz, Mary, and Kira Hall. 2004. Language and Identity. In *A Companion to Linguistic Anthropology*, ed. Alessandro Duranti. 369–94. Pastdow, Cornwall: Blackwell Publishing Ltd.

Buis, Maarten L. 2010. Direct and Indirect Effects in a Logit Model. *Stata Journal* 10 (1): 11–29. doi:10.1177/1536867X1001000104.

CEO. n.d. Centre d'Estudis d'Opinio. http://ceo.gencat.cat

Cetrà, Daniel. 2019. *Nationalism, Liberalism and Language in Catalonia and Flanders*. London: Palgrave Macmillan.

Costa-Font, Joan, and Ramon Tremosa-Balcells. 2008. Support for State Opting out and Stateless National Identity in the Basque Country. *The Journal of Socio-Economics* 37 (6): 2464–77. doi:10.1016/j.socec.2007.08.002.

Cuadras-Morató, Xavier. 2016. Catalonia: A New Independent State in Europe? In *Catalonia: A New Independent State in Europe?* ed. Xavier Cuadras-Morató. New York: Routledge.

Erikson, Robert S. 2015. Income Inequality and Policy Responsiveness. *Annual Review of Political Science* 18 (1): 11–29. doi:10.1146/annurev-polisci-020614-094706.

Galais, Carol, and Ivan Serrano. 2020. The Effects of Regional Attachment on Ideological Self - Placement: A Comparative Approach. *Comparative European Politics* 18: 487–509. doi:10.1057/s41295-019-00196-z.

Gilroy, Paul. 1994. *The Black Atlantic: Modernity and Double Consciousness*. London: Verso.

Griffiths, Ryan D. 2016. *Age of Secession: The International and Domestic Determinants of State Birth*. Cambridge: Cambridge University Press.

Griffiths, Ryan D., and Angely Martinez. 2020. Local Conditions and the Demand for Independence: A Dataset of Secessionist Grievances. *Nations and Nationalism* 27 (2): 580–90. doi:10.1111/nana.12667.

Guinjoan, Marc, and Sandra Bermúdez. 2020. Nested or Exclusive? The Role of Identities on Blame Attributoin during the Great Recession. *Nations and Nationalism* 26 (1): 197–220. doi:10.1111/nana.12547.

Guinjoan, Marc, and Toni Rodon. 2016a. Catalonia at the Crossroads. Analysis of the Increasing Support for Secession. In *Catalonia: A New Independent State in Europe?*, ed. Xavier Cuadras-Morató, 20–61. Oxon and New York: Routledge.

———. 2016b. A Scrutiny of the Linz-Moreno Question. *Publius: The Journal of Federalism* 46 (1): 128–42. doi:10.1093/publius/pjv031.

Hall, Stuart. 1997. The Local and the Global: Globalization and Ethnicity. In *Culture, Globalization, and the World-System: Contemporary Conditions for the Representation of Identity*, ed. Anthony D. King. 19–40. Minneapolis: University of Minnesota Press.

———. 2011. Introduction: Who Needs "Identity"? In *Questions of Cultural Identity*, eds. Stuart Hall and Paul du Gray, 1–17. London: SAGE Publications Ltd.

Hannan, Kevin. 1996. *Borders of Language and Identity in Teschen Silesia*. New York:

Peter Lang Publishing Inc.

Hobsbawm, Eric J. 1990. *Nations and Nationalism since 1780: Programme, Myth, Reality*. Cambridge: Cambridge University Press.

Hroch, Miroslav. 1985. *Social Preconditions of National Revival in Europe: A Comparative Analysis of the Social Composition of Patriotic Groups Among the Smaller European Nations*. Cambridge: Cambridge University Press.

IDESCAT. n.d. Institut d'Estadística de Catalunya. https://www.idescat.cat/

Iwamoto, Noriko. 2005. The Role of Language in Advancing Nationalism. *Bulletin of the Institute of Humanities* 38: 91–113.

Jaffe, A. 1999. *Ideologies in Action: Language Politics in Corsica*. Berlin: Mouton de Gruyter.

Jiménez, Tomás R. 2010. Affiliative Ethnic Identity: A More Elastic Link between Ethnic Ancestry and Culture. *Ethnic and Racial Studies* 33 (10): 1756–75. doi:10.1080/01419871003678551.

Kamusella, Tomasz D.I. 2001. Language as an Instrument of Nationalism in Central Europe. *Nations and Nationalism* 7 (2): 235–51. doi:10.1111/1469-8219.00014.

Keating, Michael. 1996. *Nations against the State. The New Politics of Nationalism in Quebec, Catalonia and Scotland*. New York: Palgrave Macmillan.

Keele, Luke, Randolph T. Stevenson, and Felix Elwert. 2020. The Causal Interpretation of Estimated Associations in Regression Models. *Political Science Research and Methods* 8 (1): 1–13. doi:10.1017/psrm.2019.31.

Labov, William. 1973. The Linguistic Consequences of Being a Lame. *Language in Society* 2 (1): 81–115. doi:10.1017/S0047404500000075.

Liñeira, Robert, and Daniel Cetrà. 2015. The Independence Case in Comparative Perspective. *Political Quarterly* 86 (2): 257–64. doi:10.1111/1467-923X.12157.

Liñeira, Robert, Jordi Muñoz, and Guillem Rico. 2020. Inferring Party Positions across Issue Dimensions. *Party Politics*. doi:10.1177/1354068820912653.

Lluch, Jaime. 2014. *Visions of Sovereignty. Nationalism and Accommodation in Multinational Democracies*. Philadelphia: University of Pennsylvania Press.

Margalit, Avishai, and Joseph Raz. 1990. National Self-Determination. *Journal of Philosophy* 87 (9): 439–61. doi:10.2307/2026968.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It. *American Journal of Political Science* 62 (3): 760–75. doi:10.1111/ajps.12357.

Moore, Margaret. 2001. *The Ethics of Nationalism*. Oxford and New York: Oxford University Press.

Moreno, Luis. 1995. Multiple Ethnoterritorial Concurrence in Spain. *Nationalism and Ethnic Politics* 1 (1): 11–32. doi:10.1080/13537119508428418.

Muñoz, Jordi, and Raül Tormos. 2015. Economic Expectations and Support for Secession in Catalonia: Between Causality and Rationalization. *European Political Science Review* 7 (2): 315–41. doi:10.1017/S1755773914000174.

Muro, Diego, and Martijn C. Vlaskamp. 2016. How Do Prospects of EU Membership Influence Support for Secession? A Survey Experiment in Catalonia and Scotland. *West European Politics* 39 (6): 1115–38. doi:10.1080/01402382.2016.1149993.

Nadeau, Richard, and André Blais. 1991. To Be or Not to Be Sovereignist: Quebeckers Perennial Dilemma. *Canadian Public Policy / Analyse de Politiques* 18 (1): 89–103. doi:10.2307/3551558.

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why*. New York: Basic Books.

Porta, Donatella della, and Michael Keating. 2008. How Many Approaches in the Social Sciences? An Epistemological Introduction. In Donatella della Porta and Michael

Keating, eds., *Approaches and Methodologies in the Social Sciences. A Pluralist Perspective*, 19–39. New York: Cambridge University Press.

Rico, Guillem, and M. Kent Jennings. 2012. The Intergenerational Transmission of Contending Place Identities. *Political Psychology* 33 (5): 723–42. doi:https://doi.org/10.1111/j.1467-9221.2012.00894.x.

Rodon, Toni, and Marc Guinjoan. 2018. When the Context Matters: Identity, Secession and the Spatial Dimension in Catalonia. *Political Geography* 63: 75–87. doi:10.1016/j.polgeo.2018.01.004.

———. 2021. Beaten Ballots: Political Participation Dynamics amidst Police Interventions. *Political Science Research and Methods*. Cambridge University Press, 1–18. doi:10.1017/psrm.2020.48.

Roeder, Philip. 2018. *National Secession: Persuasion and Violence in Independence Campaigns*. Ithaca: Cornell University Press.

Sarigil, Zeki, and Ekrem Karakoc. 2016. Who Supports Secession? The Determinants of Secessionist Attitudes among Turkey's Kurds. *Nations and Nationalism* 22 (2): 325–46. doi:10.1111/nana.12150.

Serrano, Ivan. 2013. Just a Matter of Identity? Support for Independence in Catalonia. *Regional & Federal Studies* 23 (5): 523–45. doi:10.1080/13597566.2013.775945.

———. 2019. Ethnic Alignment in Divided Regions: Individual and Contextual Factors. *Territory, Politics, Governance* 8 (4): 532–52. doi:10.1080/21622671.2019.1582357.

Sorens, Jason. 2005. The Cross-Sectional Determinants of Secessionism in Advanced Democracies. *Comparative Political Studies* 38 (3): 304–26. doi:10.1177/0010414004272538.

Tormos, Raül, Jordi Muñoz, and María José Maria José Hierro. 2015. Endogenous

Identities? How the Independence Debate Is Reshaping Catalans' Identity. Paper

Presented at the Fòrum Rosa Virós 2015 (Barcelona, 4th June 2015).

Verkuyten, M. 2005. *The Social Psychology of Ethnic Identity*. Hove: Psychology Press.

Webster, David. 2007. From Sabang to Merauke: Nationalist Secession Movements in

Indonesia. *Asia Pacific Viewpoint* 48 (1): 85–98. doi:10.1111/j.1467-

8373.2007.00332.x.

NOTES

[1] It is beyond the scope of this research to determine why certain groups develop strong regional feelings that do not eventually develop into secessionist claims. For this topic, see Roeder (2018) or Lluch (2014).

[2] Several aspects need to be weighted up concerning the ancestry variable. First, the use of a three-fold variable considering the parents' origin allows better identifying the mechanism at play than a simple dichotomous measure tapping into the respondent's origin (Catalonia or not). This is so because up to 98.3 per cent of the respondents with the two parents born in Catalonia are themselves born in the region—the percentage goes down to 92 per cent among those respondents with one parent born in Catalonia, and to 37.3 per cent among those with none of the parents born in Catalonia. Hence, de facto, the ancestry variable incorporates embedded information on the individual's origin—via the category of response 1–while allowing for better nuance in the remaining two categories (0 and 0.5). Second, a test for linearity reveals that this 3-categories variable is linearly related to support for secession. In other words, percentage of support for secession among those individuals with one parent born in Catalonia (value 0.5) is exactly halfway the values 0 and 1—evidence available in the Replication materials. This empirical phenomenon allows me to use a single measure of ancestry instead of coding it as a series of dichotomous variables—and thus potentially making the analysis and the interpretation of results more intricated.

[3] The inclusion of other control variables such as income, political knowledge or party identification provide the same empirical results.

[4] The results remain unaltered if I do not include the causal link between ancestry and language.

[5] Despite data from the CEO is available before 2014, there are several reasons that recommend focusing on the 2015-2020 period. First, in 2014 the CEO changed the survey methodology, from phone survey to a face-to-face survey. Remarkably, not only the survey methodology was changed but also the weights of the different social groups in society were reconsidered in the survey design. Hence, while in surveys performed before the 2014 the percentage of Catalan

and Spanish speakers was often evenly distributed, in surveys performed since 2014 the percentage of Catalan and Spanish speakers already matches actual figures in Catalonia (36 percent Catalan, 48 percent Spanish). Thus, among the reasons that explain why starting in 2014 support for secession and the percentage of people feeling exclusive Catalan decreased in surveys—likely, together with a genuine reduction for political reasons—was the change in methodology and the new weights assigned to the different social groups in Catalonia. Second, my dependent variable, a dichotomous measure of support for secession, was not included in the CEO surveys until February 2015—note, therefore, that the 2014 face-to-face surveys do not include such question. Before 2015, support for secession was only indirectly identifiable through a question tapping into the preferred relationship between Spain and Catalonia (a region, the status quo, federalism or independence). Third, as mentioned above, support for secession and national-regional identity have remained considerably more stable during this period than during the phone-survey period (2006-2013).

[6] Replication files can be found at Harvard Dataverse (https://doi.org/10.7910/DVN/NEAPLH). The different barometers were administered in February, June and October 2015; February, June and October 2016; March, June and October 2017; April, June and October 2018; March, July and November 2019; and February 2020.