Capturing Ethnicity: The Case Of Ukraine

Olga A. Onuch\(^a\)* and Henry E. Hale\(^b\)

\(^a\)Politics Department, University of Manchester, Manchester, UK; \(^b\)Elliott School of International Affairs, George Washington University, Washington, DC, USA

(Received 24 December 2017; accepted 25 January 2018, finalized February 22, 2018)

*Corresponding author. Email: olga.onuch@manchester.ac.uk

Building on past survey-based studies of ethnic identity, we employ the case of Ukraine to demonstrate the importance of taking seriously the multidimensionality of ethnicity, even in a country that is regarded as deeply divided. Drawing on relational theory, we identify four dimensions of ethnicity that are each important in distinctive ways in Ukraine: individual language preference, language embeddedness, ethnolinguistic identity, and nationality. Using original survey data collected in May 2014, we show that the choice of one over the other can be highly consequential for the conclusions one draws about ethnicity’s role in shaping attitudes (e.g., to NATO membership), actions (e.g., participation in the Euromaidan protests), and the anticipation of outgroups’ behavior (e.g., expectations of a Russian invasion). Moreover, we call attention to the importance of including the right control variables for precisely interpreting any posited effects of ethnicity, making specific recommendations for future survey research on ethnic identity in Ukraine.

Keywords: ethnicity; identity; language; Ukraine; public opinion; political behavior
Introduction

How does ethnicity influence politics in Ukraine? Readers familiar with the volumes of literature on post-Soviet Ukrainian identity might be forgiven for considering this question to have already been resolved. Ukraine, the common wisdom goes, is a divided society, and this divide shapes everything from policy attitudes to protest participation. Yet once one looks more closely at the purported identity divide, its boundaries start to become fuzzier, even to the point at which it is unclear precisely who is a true “Ukrainian” and who is not. In fact, most of the country’s residents do not perceive sharp cultural distinctions between “Ukrainians” and even their purportedly chief ethnic “other,” “Russians.” Accordingly, if one drills down deep into the actual findings presented by studies of Ukrainian ethnic politics, one finds broad agreement only on a basic notion that ethnicity often matters, with little broadly accepted knowledge as to how, why, and when it matters.

This state of affairs is likely to reflect (at least in part) a problem of measurement. With the thorough debunking of primordialist notions of ethnicity—whereby ethnic groups are considered unproblematically to be clear cut, unchanging over modern history, and naturally salient, hence easy to identify—cutting-edge research on ethnic politics has wrestled with how to measure ethnic categories that can be blurry, shifting, overlapping, and permeable (Chandra 2012). Indeed, how can one hope to ascertain the effects of an ethnic divide on politics if one cannot clearly measure where one group begins and the other ends? In the last two decades, most of the advances in measurement have come at the macro level, with new datasets being created that avoid primordial assumptions while still identifying the most salient ethnic groups in a country and generating useful statistics that characterize these groups (Fearon 2003; Wilkinson 2004; Posner 2004; Chandra and Wilkinson 2008; Cederman, Weidmann, and Gleditsch 2011; Marquardt and Herrera 2015). Much less attention has been paid to how we collect data at the micro level, how we uncover the ethnicity of a given individual, and it is here that the clarity of even the best new macro-level datasets starts to dissipate. As a result, most micro-level studies of ethnicity that attempt to aggregate their findings wind up rather arbitrarily selecting indicators of an individual’s ethnicity, without the benefit of careful, prior, theoretically informed analysis of what precisely each available indicator actually represents.

In the following pages, we draw on original survey data from May 2014 to demonstrate that we face a major problem of measurement in attempting to understand ethnic politics in Ukraine. In particular, we show how different commonly used measures of Ukrainian ethnicity can make the difference between “finding” that ethnicity matters and “finding” that it does not matter for important political outcomes such as supporting NATO membership, expecting a Russian
invasion, and turning up on the streets with the Euromaidan protest movement. Drawing from a relational approach to understanding the nature of ethnicity, we argue that Ukrainian ethnicity is characterized by at least four distinct dimensions that only imperfectly overlap and that have distinct effects. We label these dimensions individual language preference, language embeddedness, ethnolinguistic identity, and nationality. For the most precise interpretations of the nature of ethnicity’s effects on political outcomes using regression analysis, we recommend that measures of all four be included in our models, either as controls or as quantities of interest. We conclude that in the case of Ukraine, ethnicity’s effects on political attitudes tend mainly to reflect personal language preferences and identification with different notions of Ukrainianness, while ethnicity’s facilitation of joining the Euromaidan operated primarily through the networked mobilization of Ukrainian-speaking workplaces. Ethnicity appears not to have shaped Ukrainians’ expectations of a Russian invasion.

**Ukraine and the multidimensional nature of ethnicity**

Identity can be fruitfully defined as the “set of points of personal reference on which people rely to navigate the social world they inhabit,” with ethnic identity (or its synonym “ethnicity”) referring to one subset of those points of personal reference (Hale 2008, 34, 40). That is, identity is essentially a cognitive uncertainty-reducing mechanism, a process by which the brain copes with the vast complexity of the social environment by breaking it down into meaningful categories (Gaertner et al. 2002). Shared reference points (such as having a particular physical trait in common) can come to define identity categories or groups, and such categories or groups can take on greater meaning for the individual to the extent that the reference points in question potentially have large implications for one’s life chances.

Chandra (2012, 9) has made a powerful argument that what distinguishes ethnic categories from others is that “descent-based attributes are necessary for membership.” To say a category is descent-based is to say neither that it is genetic and unchangeable nor that descent must be the only criterion for membership. Rather, it is to say that membership in a given category is obtained through some kind of mechanism that typically depends heavily on descent. This definition of ethnicity thus includes categories based on shared physical or “racial” features, which are commonly inherited genetically, as well as categories for which membership depends on language (usually passed along through families), broader sets of cultural attributes (also passed along in part through families), or the nationality of one’s parents (2012, Chapter 2).

The power of ethnicity relative to other types of identity comes from several features of descent-based categories that make them useful to people as rules of
thumb for understanding their constraints and opportunities in life. Chandra (2012) observes that the descent-based attributes that define ethnic categories tend to be readily perceptible (e.g., the visibility of skin color or the audibility of language or accent) and also hard to change. For example, a Russian-speaker in Donetsk can instantly detect when Ukrainian is being used instead of Russian, and it can be hard for people raised in an exclusively Ukrainophone neighborhood by exclusively Ukrainophone parents to learn the Donetsk dialect of Russian so perfectly as not to be noticed.

Moreover, Hale (2008, Chapter 3) argues that ethnic (descent-based) categories also tend to overlap with other important features of the social world. This means, for one thing, that readily perceptible, hard-to-change ethnic traits can become easy shorthands for inferring a lot of information about a person one does not actually know. For example, prior to 20th century upheavals like Soviet industrialization and World War II, the speaking of Ukrainian was most common in rural areas, with big cities more populated by those speaking Polish or Russian. The overlap between speaking Ukrainian and rural origins led many in Ukraine to associate the Ukrainian language with rural life and its related behaviors, beliefs, and socioeconomic status. The hearing of Ukrainian thus still evokes associations of low social status and rural culture among some residents of Ukraine or other parts of the former Soviet Union. But since the demise of the USSR and the rise of a “Ukrainian” state, speaking Ukrainian has become more associated with political power and dominant social status within Ukraine (Bilaniuk 2005, 15). Moreover, the overlap of ethnicity with the distribution of other factors impacting life chances can encourage a sense of linked fate with others who share membership (Hechter 1975; Dawson 1994). Thus when ethnocentric Russian politicians make condescending statements about Ukrainians’ accents, the affected Ukrainophones feel that they are impacted in a similar way, their fates intertwined. The same is true for Russophones when extreme Ukrainian nationalists call for banning the Russian language in key spheres of life. The degree to which each dimension of identity conveys associations like these that people consider important for their life chances constitutes the degree of *thickness* of that identity dimension (Geertz 1973, 3–30; Hale 2008, 36).

The extent to which a country’s thickest dimensions of ethnicity tend to overlap or cut across one another can powerfully influence how people understand politics in that country. When the thickest ethnic divisions tend to overlap more than they cross-cut, they can form what scholars call a divided society in which ethnic cleavages tend to subsume other interests in structuring political competition (Horowitz 1993; Chandra 2005). Scholars inside and outside Ukraine frequently regard it as just such a case, one in which the chief cleavages overlap more than cross-cut in producing what is often described as an east-west or a Europhilic-Russophilic divide (Riabchuk 1992, 2002; Hesli, Reisinger, and Miller 1998; Walker
Crucially, however, it is well nigh impossible for different dimensions of ethnicity in a country to overlap perfectly, meaning that we must be attentive to the potential for different thick identity dimensions to denote different things even when their association is high. We see this quite clearly in the case of Ukraine, where at least three dimensions of ethnic identity are argued to be thick enough to be a driving force behind political competition in Ukraine but do not align perfectly (Shevel 2002). First, many accounts portray a Ukraine cleaved primarily by language use, with politics revolving most fundamentally around a competition between Ukrainophones and Russophones (Arel 1995; Laitin 1998; Bilaniuk 2005; Wolczuk 2006; Colton 2011; Kulyk 2011, 2013; Charnysh 2013). Other studies describe the most essential ethnic divide as being centered around self-declared nationality, in particular whether one identifies mainly with the categories of “Ukrainian,” “Russian,” “Crimean Tatar,” or something else. This cleavage, based more on historically developed ethnic self-consciousness than language or any other specific ethnic trait, is said to have developed through some combination of pre-communist schooling, enduring family traditions, Habsburg and Soviet nationalities policies, and mobilization by ethnic activists (Beissinger 1988; Arel 1993; Bremmer 1994; Pirie 1996; Dawson 1997; Kulyk 2001; Darden and Grzymala-Busse 2006; Hale 2008; Bard 2014; Blum 2014). Third, Kulyk (2011) has advanced a compelling argument that “language identity” can in fact constitute an identity dimension distinct from that based on actual language use (practice). His insight is that many people in Ukraine identify with the Ukrainian language even though they do not actually speak it and/or even though they might not be considered “Ukrainian” by nationality, and that this identification may be the most important ethnic driver of attitudes in Ukraine.

In cases such as this where cleavages defined by different dimensions of ethnicity overlap but do so imperfectly, there is likely to be competition over what the “essential” boundary markers separating one group from another are and hence what the “real” nature of the group is (Wimmer 2013). This is because individuals can have very different interests in where exactly the boundaries are drawn and enforced (Barth 1969). The self-interested contestation of group boundaries can be observed readily in Ukraine. For example, President Leonid Kravchuk, fighting a losing battle to retain office in 1994, famously insinuated that because challenger Leonid Kuchma was not a Ukrainian speaker he not going to defend the Ukrainian nation from Russian interests. By implication, a vote for Kravchuk was a vote for an “independent Ukraine” (Kolomayets 1994). Kuchma, for his part, de-emphasized language even as he titled his memoirs “Ukraine is not Russia” (Kuchma 2004). More recently, the well known novelist Yurii Andrukhovych sparked controversy by criticizing the use of Russian language terms in western Ukraine (Gazeta.ua 2017). Analogously, high-ranking members of the Orthodox Church in Ukraine that
recognizes the Moscow Patriarchate have made calls for “religious unity” under “one local church,” essentially putting the church at the center of what it means to be Ukrainian in a way that would link it to Moscow (Pravoslavie.ru 2017). In short, each individual is likely to have a slightly different relationship to each category, creating space for a powerful politics of ethnic boundary-making that many regard as central to ethnic politics worldwide (e.g., Wilkinson 2004; Laitin 2007; Wimmer 2013).

For this reason, it is important not to go into a study of ethnic politics in any one country presuming the existence of well-defined, clearly delineated “ethnic groups” (Brubaker 2004). Instead, it is likely to be much more productive to think not in terms of groups but in terms of distinct ethnic categories (dimensions of ethnicity), the convergence or overlap of which should be problematized and researched rather than assumed. And researchers should be very sensitive to the possibility that different dimensions of what is sometimes regarded as the same “ethnic group” might in fact influence people in different ways (that is, people may differentially identify with different ethnic markers). Unfortunately, as Chandra and Wilkinson (2008, 520) once pointed out, social scientists attempting to study ethnicity using quantitative methods too often select measures of “ethnicity” rather arbitrarily (including using whatever measure just happens to be available), sometimes appearing to do so without questioning “how the data were generated and what they mean,” leading to “ad hoc” interpretation. The study of Ukraine turns out to be no exception.

Untangling ethnicity’s effects in Ukraine using survey data

While myriad scholarly works address the origins and nature of Ukrainian identity, surprisingly few have attempted rigorously to assess which dimensions of ethnicity are most important using survey data.¹ Survey research would seem to be particularly useful for this endeavor because samples can be designed to be representative of the country as a whole (providing strong grounding for claims about nationwide patterns) and because the resulting data reflect what Ukrainians themselves actually say in response to questions specifically designed to study ethnicity and its effects.² The pioneering work has been done, however, by only a handful of scholars, perhaps most prominently Kulyk (Kulyk 2011, 2013), Barrington and his co-authors (Barrington 2002; Barrington and Herron 2004; Barrington and Faranda 2009), and Shulman (2005).

Much remains to be done, however. First, most of the pioneering studies have been more focused on sorting out the effects of “region” from those of ethnicity than they have been on disentangling different dimensions of ethnicity from each other. The conclusion has consistently been that living in a given region has distinct implications on behavior that cannot be accounted for by ethnic difference (Birch
2000; Kubicek 2000; Barrington 2002; Barrington and Herron 2004; Clem and Craumer 2005; Barrington and Faranda 2009; D’Anieri 2011; Osipian and Osipian 2012). Since this finding is now rather uncontroversial, our study does not enjoin the “regions” debate in Ukraine, instead following Kulyk (2011) in training the microscope more directly on the possible differential effects of different dimensions of ethnicity. We do, though, adopt a key recommendation from studies of regional effects by controlling for spatial patterns when estimating ethnicity’s effects.

Second, the most prominent trailblazing survey-based studies of ethnicity in Ukraine, while including both “language use” and “nationality” in their analyses, tend to treat these as unproblematic categories. They thus typically include only a single indicator for each with little or no discussion of how the chosen measure is constructed precisely, how one possible measure might differ from another, or (consequently) how pairing the “nationality” indicator with different indicators of “language” (that is, controlling for slightly different things) might yield different findings on each variable. For example, while Barrington and Faranda (2009, 249) find that living in Ukraine’s east changes the effect of being a Ukrainian speaker, they do not consider how their findings on nationality might be influenced by their choice of indicator for “language.”

Third, the specific measures used to capture “language” vary from study to study. This does not apply to the concept of nationality: the trailblazing works virtually all use a measure essentially inherited from Soviet-era censuses that asks people their “nationality” and gives them a choice of Russian, Ukrainian, or “other” (and no option for mixed-nation identity). But the measures these studies use to capture “language” differ greatly across studies. To take some of the most prominent works as examples, Barrington (2002) codes respondents’ language as the one that the interviewer noticed them mainly using during the survey, Barrington and Faranda (2009) and Shulman (2005) use the language that respondents say they speak most often at home, while Kulyk (2011) employs both a measure of usual language practice (described as the language of “everyday use”) and an indicator capturing the language respondents say is their native tongue.

The survey-based literature has thus not yet wrestled with the full complexity of ethnicity in Ukraine, in particular the imperfect overlap among different dimensions of ethnicity with distinct and important implications for individuals’ real and perceived life chances. One result has been that findings have failed to accumulate beyond the general conclusions that (a) ethnicity matters in Ukraine; and (b) how ethnicity matters there is complex, with language and nationality mattering at some times but not others depending on study design. Research by Kulyk (2011, 2018) has shed new light on the meaning and influence of “native language” and “nationality,” but these important findings have not been adequately engaged by other scholars, and their relationship to other language-
related dimensions of ethnicity remains largely unexplored. Table 1 provides a brief summary of this state of affairs with reference to some of the most prominent works that explicitly set out to distinguish among the effects of different dimensions of ethnicity in Ukraine.

|TABLE 1 ABOUT HERE|

**A relational approach**

A good starting point for more fully engaging Ukraine’s ethnic complexity is theory. What would the understanding of ethnicity as certain descent-related points of personal reference (as discussed above) lead us to expect would be the relationship among different, imperfectly overlapping dimensions of identity in a case like Ukraine? First and foremost, we would anticipate that each such dimension would reflect a distinct relationship between that individual and the rest of the world. At the same time, however, Ukrainians may be using each dimension as a shorthand (or rule of thumb) for discerning other information about people that may be systematically correlated with that particular dimension of ethnicity (or perceived to be). That is, each dimension may hold meaning beyond what is intrinsic to it. And from a practical standpoint, to the extent that dimensions of ethnicity in Ukraine are correlated with each other, the meaning that is specific to each dimension may not clearly emerge in our analyses without controlling for other dimensions of ethnicity. From this perspective, neither language use nor nationality (the two aspects of ethnicity widely deemed most important in Ukraine) can be interpreted as straightforward, uncomplicated categories of analysis.

Let us first take the Ukrainian language. While prior studies usually claim simply to identify “people who speak Ukrainian” and to study the effects of speaking that language, we argue for considering at least three different ways in which language can function as an ethnic point of personal reference with distinct implications for behavior on the part of the individual in question. First, we might identify the language that an individual would choose to speak if communication partners were all completely indifferent and no social desirability considerations were in play. We will call this category *individual language preference*. This reflects an individual’s deepest personal relationship to a language, including “comforts of home” that a person might experience in using it instead of another for communication (Hardin 1995, 89; Laitin 2007, 56).

This is not what is most directly captured by most of the indicators used in studies of language identity in Ukraine, however. Three of the four prominent studies summarized in Table 1, for example, rely on measures of the language people primarily speak “at home” or “every day,” with other variations specifying the language used “in private life” or “at work.” But these language practices are likely
to be strongly influenced by the language repertoires and language choices of one’s spouse, friends, acquaintances, or co-workers, and thus might not in fact reflect the individual’s own, personal language preference (not least because we know that generational, gender, and other hierarchies and power dynamics exist and inform our social exchanges). Instead, what we think these measures uniquely capture is what we call language embeddedness, or language use made under the influence of social environments in which one is embedded. Being embedded in a Ukrainian-speaking language environment regardless of what language one might actually prefer to speak is likely to be associated with interests or viewpoints that may be shared by or conveyed through Ukrainian-speaking networks. And since language structures thinking to a significant degree, embeddedness in Ukrainian-language settings is also likely to foster adherence to any belief structures or behavioral norms that may be intrinsic to the Ukrainian language itself and thus cognitively activated when using the language, a possibility that linguists have documented (Laitin 1977; Laitin 2007, 72–73).

How can we sort out the effects of language embeddedness from individual language preference in our survey analyses? The key is to come up with a measure that is explicitly designed to minimize the social pressures involved in expressing a language preference. Perhaps most prominently, the Kyiv International Institute of Sociology (KIIS) has developed and regularly uses the following protocol for its nationally representative surveys in Ukraine: a bilingual interviewer begins the interview with a greeting that sounds the same in Ukrainian and Russian, notes the language in which the respondent replies, follows up by asking in that language whether the respondent is more comfortable speaking that language or another in the interview, and then records the choice of language ultimately made by the respondent. While respondents sensitive to social desirability considerations are still likely to guess which language the interviewer might prefer and adapt their responses accordingly, this social pressure is likely to be significantly lower than in situations like “work” or “private life” where interlocutors do not systematically attempt to minimize them. A related measure, also employed by KIIS in its standard protocol, has the interviewer observe, during the course of the entire interview, what language (or mix of languages) the respondent actually spoke. We would expect this to be a less direct indicator of individual language preference since as the interview progresses, the respondent will have more information about the interviewer that may invoke social desirability considerations. Nevertheless, while no survey measure is perfect, it is important to distinguish among them when designing and interpreting survey results regarding the “effects of language.”

Further complicating the study of “language effects” in Ukraine, asking people for their “native language” or “mother tongue” will yield results that differ significantly from either individual language preference or language
embeddedness (Aza 1995; Arel 2002). Specifically, when asked to give their “native language,” many people reply “Ukrainian” when their primary languages of both preference and use are Russian, and even when they do not speak Ukrainian at all. In his pioneering study, Kulyk (2011) calls this phenomenon “language identity,” the origins of which are as follows. The conception of ethnic identity constructed by the USSR put “native language” at the center, with each major group (e.g., Ukrainian, Kazakh, Georgian) having its own territory (in Ukraine’s case, a union republic) and also an official “native language” that was treated as a crucial boundary marker between groups (Slezkine 1994; Martin 2001). When the USSR dissolved, the Ukrainian state adopted a civic rather than an ethnic definition of what it means to be “Ukrainian” (Brubaker 1996, 19), effectively creating two notions of “Ukrainianness”: a narrower ethnic conception with “native language” at the center and a civic notion that is much more about identification with the state than with any particular group within it (Shevel 2002). In this sense, what is being picked up in questions about one’s “native language” is essentially adherence to the older ethnic notion of what it means to be Ukrainian, with the associated language being a matter of descent-based identification rather than actual practical use (Kulyk 2018 [Query 3]). Since this form of identification is not primarily about language itself but about identifying with what is seen to be a much more robust and historically developed ethnic category, we recommend referring to it conceptually not as “language identity” but as “ethnolinguistic identity” in order to make clear this broader meaning.

This distinction between Ukrainian ethnolinguistic identity (measured by asking about native language) and civic identification with the Ukrainian state is also very important for interpreting answers to survey questions designed to elicit nationality. When individuals are asked what their “nationality” is and answer “Ukrainian,” the convention is to assume that respondents are referring to a clear-cut Soviet-census-style ethnic category that is clearly distinct from other categories such as “Russian,” “Georgian,” or “Belarusian” and that implies identification with a whole cluster of ethnic traits that Ukrainians are believed to share. But with Ukrainian internal passports no longer declaring an individual’s nationality according to this Soviet definition, and with the “Ukrainian state” declaring a civic definition of citizenship and a broader civic “nation,” the term “nationality” has increasingly come to refer to an individual’s relationship to the state and is increasingly connoting identification with a more inclusive, broader notion of what it means to be Ukrainian (Shevel 2002; Sasse 2010; Kulyk 2013). Research has thus found that when we ask individuals to report their “nationality” in surveys, some people report being “Ukrainian” who do not strongly identify with the Ukrainian language or even other traditional markers of Ukrainianness, including people formerly identifying as “Russian” (Kulyk 2018).

Even beyond this rethinking of the meaning of self-declared nationality in Ukraine,
a relational approach to ethnicity would also lead us to problematize rather than assume “either-or” definitions of nationality because people can identify with multiple categories at the same time and to different degrees with different associations. To assume Ukrainian and Russian nationality are mutually exclusive can have at least two implications. First, we may overlook the possibility that people of “dual nationality” might behave distinctly in ways that do not simply reflect a mid-point between how “Ukrainians” and “Russians” behave. Due to space limitations and our own concern here specifically with dimensions of Ukrainian identity, we leave this question for future research. The implication that we do address here is that behavior might vary according to the degrees of self-identification with the category “Ukrainian” relative to “Russian,” with the category “Ukrainian” having more meaning for the individual and thus being more strongly influential on behavior the greater the proportion of “Ukrainianness” one feels. We propose that in this case, a scale of “Ukrainianness” might more closely correlate with the distribution of attitudes and behaviors than would a binary, forced-choice indicator of nationality. On the other hand, an axiom of survey research is that forced choices frequently reveal meaningful preferences by eliminating an “easy middle” option for escaping revealing that preference (Schuman and Presser 1996; Dhar and Simonson 2003). If this were true, we would expect to find the binary measure of Ukrainian nationality to be a stronger predictor of behavior than a scale, which would “muddy” the measure.

What, then, are implications for how we should construct survey-based studies of ethnicity’s effects in Ukraine? At the most general level, we would expect all of these dimensions of Ukrainian ethnic identity to be rather strongly correlated with one another at the same time that we expect the correlation to be imperfect. That is, there is likely to be a large cluster of individuals who measure “positively” on most or all of these variables, driving high correlation coefficients. And because this cluster of people is large, including any one indicator by itself as an explanatory variable in an econometric equation (that is, without adding other ethnic controls) means that this indicator is likely to “represent” not only itself but all of the other dimensions of Ukrainian ethnicity that are correlated with it. If this single representative “Ukrainian ethnicity” variable is statistically significant, we can conclude that there exists some sense in which “Ukrainian ethnicity” is correlated with the outcome of interest.

At the same time, to the extent that there are also many people for whom these dimensions of ethnicity do not perfectly align, which we hypothesize is frequently the case in Ukraine, then the choice of which single representative “Ukrainian ethnicity variable” to include in the regression will matter a great deal for interpreting the results in at least two ways. First, if this variable is significant, we will not be able to tell for sure whether the detected effect owes to a logic of individual language preference, language embeddedness, ethnolinguistic identity,
or civic national identity. That is, even if one’s chosen single variable is (say) individual language preference and it is statistically significant, we cannot conclude that the detected effect is in fact about individual language preference since it is correlated with and thus likely to be “picking up” the effects of (say) ethnolinguistic identity in the analysis. Beyond follow-up interviews with respondents, the simplest way to identify whether the effect really is due to individual language preference would be to include ethnolinguistic identity as a control variable. Control variables are thus crucial for precise interpretation of ethnicity’s effects.

Second, because different dimensions of Ukrainian ethnicity can be expected to connect the respondent with the outcome of interest in a distinct way, there is a chance that the choice of single indicator could determine whether or not the hypothesis is confirmed that Ukrainian ethnicity in general has significant effects. And there is a danger of “data mining” here: since a 95% statistical significance standard essentially means one can expect a false positive one in twenty times, arbitrarily “trying out” different single indicators of “Ukrainian ethnicity” to test for its effects increases the risk of encountering one of these false positives and mistakenly concluding ethnicity has an effect when in fact it does not.5

Since most studies of ethnicity posit or are interested in identifying specific mechanisms by which ethnicity has effects, and in order to minimize the dangers of data mining, we argue it will usually be prudent to include indicators of four distinct dimensions of ethnicity in statistical analyses of Ukrainian ethnicity’s effects using survey data.6 These dimensions (or “attribute-dimensions” in Chandra’s (2012) terminology) are those that prior research has argued do tend to be important in Ukraine in different ways: (a) individual language preference, (b) language embeddedness, (c) ethnolinguistic identity, and (d) civic nationality. If we do this, we can be confident that when “native language” is significant, for example, the detected effect indeed has to do with ethnolinguistic Ukrainian identity as opposed to individual language preference, language embeddedness, or identification with the Ukrainian state.7

Data, operationalization, and method

To demonstrate the importance of considering the different meanings of different measures of Ukrainian ethnicity, we draw on results from an original survey that measures each of these four dimensions of ethnicity and, to our knowledge uniquely, includes multiple measures of each where possible. We first assess the degree to which these measures are correlated with each other, demonstrating the validity of our four-category characterization of ethnicity in Ukraine, and then show that different measures and different combinations of measures have different relationships to important attitudes, expectations, and behavior in the
ways anticipated by our theory.

The empirical heart of our study consists of original survey data collected by the authors in collaboration with KIIS during 16–24 May 2014, right before Ukraine’s 25 May presidential election. A nationally\textsuperscript{8} representative sample of 2,015 individuals was selected through a stratified, multi-stage, area probability technique, with a 51% response rate.\textsuperscript{9} The survey fully included the Donbas with relatively minor substitutions of statistically equivalent communities for zones affected by the conflict, which was only emerging at that time. The margin of error of our frequency estimates is no greater than 3.3%.

Our key independent variables of interest are measures of each of the four dimensions of ethnicity that we argue belong in analyses of Ukrainian ethnicity’s effects:

1. \textit{Personal language preference}. Initially, we record simply whether the respondent, at the outset, chose to answer the questionnaire in Russian or Ukrainian after the interviewer initially administered the standard KIIS introductory procedure described above, creating a binary variable indicating whether they took the survey in Ukrainian. The language initially chosen, however, was not necessarily the language in which the respondent continued to answer in practice. Thus we also create a binary variable capturing whether, in the interviewer’s judgment, the respondent used mostly Ukrainian during the interview.\textsuperscript{10}

2. \textit{Language embeddedness}. We asked respondents to tell us the language they mainly speak in their private lives as well as the tongue they primarily use at their place of employment. We code these as two binary variables, one for people speaking mostly Ukrainian in their private lives and the other for those speaking predominantly Ukrainian at work.\textsuperscript{11} We expect these to capture embeddedness in Ukrainian-speaking social environments.

3. \textit{Ethnolinguistic identity}. We examine results from a standard question asking people to report their mother (native) tongue (as practiced by Ukrainian census takers), creating a binary variable coded 1 for people who select Ukrainian.\textsuperscript{12}

4. \textit{National identity}. Here we consider two measures. One asks people the degree to which they consider themselves to be Ukrainian, yielding a five-point scale of “Ukrainianness” that ranges from “entirely” to “none.” \textsuperscript{13} We then create a “forced choice” measure that records the
answers when people are then required to choose the single category with which they most strongly identify.¹⁴

To demonstrate the importance of making such distinctions when using or interpreting survey data on ethnicity’s effects, we consider three types of effects that the literature suggests ethnicity can have: effects on attitudes (Darden and Grzymala-Busse 2006), effects on expectations regarding what other people will do (Kaufman 1996; Hale 2008), and effects on actual behavior (Beissinger 2013). Moreover, we “weight the dice” against our expected finding that using different measures of ethnicity would lead to different conclusions by studying phenomena that experts widely regard as straightforwardly and robustly driven by Ukraine’s ethnic divide.

1. Language policy: whether people support Russian becoming an official state language in Ukraine,¹⁵ coded as a binary variable. Our survey finds 46% of the population to be generally supportive.

2. NATO membership: whether people support Ukraine’s joining NATO,¹⁶ also coded as a binary variable; 31% backed this proposition according to our data.

3. Expectations of a Russian invasion in 2014: whether the respondent felt (following warnings by many Ukrainian politicians) that it was likely Russia would invade Ukraine as of May 2014,¹⁷ coded as a binary variable. In May 2014, we find that 36% thought this at least somewhat likely.

4. Participation in the Euromaidan protest movement: a binary variable capturing self-reported participation in the Euromaidan.¹⁸ We register 10% of the population participating either in Kyiv or a local part of the protest movement.

Since we are interested in the full effects of ethnicity, including effects that are likely to be mediated by other variables (such as adherence to nationalist ideology), our analysis needs only to control for factors that are likely significantly to influence both ethnic identification and our dependent variables of interest. Most obviously, therefore, we adopt standard demographic control variables in case patterns of ethnic identification are influenced by gender, age, education,¹⁹ and residence in an urban environment (a populated point of at least 50,000 residents). In addition, for reasons given above, we also control for region, adopting the widely used practice of dividing Ukraine into four categories (east, west, south, and center), including three in our models and treating the fourth (west) as the reference category.²⁰ Finally, we include controls for individuals’ economic
condition, which some studies find can shape identity (Weber 1978; Hechter 2000): a binary variable capturing people who are “transition winners” relative to 1991 (Tucker 2006) and a seven-point scale measuring the respondent’s family’s financial situation. Our survey’s estimates of the distribution of these dispositions in Ukraine can be found in an online appendix on the Harvard Dataverse, which also includes a replication dataset.

Since our dependent variables are all binary, we employ logistic regression models to estimate the effects of our ethnic variables. Because logit coefficients do not facilitate straightforward interpretation, in what follows we report the estimated effects of each factor on a given dependent variable in terms of full effects. A full effect is the average marginal effect when all variables are scaled from 0 to 1. Or put otherwise, it is the average change it makes in an individual’s estimated likelihood of participating in the EuroMaidan, believing Russia will invade Ukraine, having a preference for NATO membership, or having a preference that Russian becomes an official state language, when one raises a given factor from its minimum to its maximum value in the dataset while holding all other variables at their actual values in the dataset. Full effects express a clear idea of the full range of variation that a given variable is found to produce in a manner that facilitates comparison with other variables’ effects, thus making them expedient for reporting results and discussing implications.

Results

To begin, we examine patterns of correlation within the set of all seven indicators of Ukrainian ethnicity that we consider in our study. Table 2 presents Pearson correlation coefficients for each pair of these measures, showing an overall high degree of intercorrelation that ranges from a minimum of 0.222 for our binary measure of nationality and language used at work to a maximum of 0.803 for the language an individual uses in private life and the language a respondent is observed using to take the survey. Each indicator is thus to some degree correlated with the others.

At the same time, the specific patterns reported in Table 2 do appear to justify our theory-driven decision to break these measures down into four distinct dimensions of Ukrainian ethnicity. To begin, the correlations among these variables are far from perfect, indicating that the boundaries separating ethnic groups in Ukraine are not in full alignment, leaving room for contestation and differential effects of identifying with one particular boundary marker over another. Moreover, the highest correlation coefficients generally occur among variables that are part of the same dimension of Ukrainian ethnicity as we have defined them. Our two
indicators of personal language preference (survey language chosen and survey language observed) are correlated with each other at a level of 0.76, reflecting a tighter mutual relationship than either of these measures has with any other measure (with one exception discussed below). Similarly, the indicators of language embeddedness (language of private life and language of work) are intercorrelated at a level of 0.66, more than with any other measure of ethnicity (with the same exception just mentioned). And the binary and gradated measures of nationality are mutually correlated at 0.78, a closer relationship than either measure has with any other ethnic indicator. The only exception, an instance where the correlation between two variables crosses the line between our dimensions of Ukrainian ethnicity, is the close (0.80) relationship between language of private life and language observed in the survey.  

This multidimensionality of the boundary separating “Ukrainians” from other groups in Ukraine coincides with a sense that Ukrainians are really not very different from Russians, the largest other putative group in the country. Our survey’s respondents were asked to use an 11-point scale to assess the degree to which “Russians” and “Ukrainians” are essentially identical to or completely different from each other in terms of values and behavior. We find that only 2% of the population replied that they were completely different, and a total of just 11% chose any response on the “more different” side of the scale. Moreover, only 11% chose the middle response, leaving an overwhelming 72% of citizens to reply clearly that Russians and Ukrainians are more similar than different (the leftover 6% percent did not provide a substantive answer).

Which dimensions of Ukrainian identity are most associated with perceiving a thicker divide between Ukrainians and Russians? Figure 1 reports the results of a tobit regression analysis designed to test whether each of our four primary dimensions of Ukrainian identity (here operationalized by the respondent’s chosen survey language, language of private life, native language, and forced-choice nationality) are significant predictors of a thicker perceived boundary between Ukrainians and Russians, coding the responses on the scale discussed above such that higher values represent perceptions of a more pronounced Russian-Ukrainian divide.  

It turns out that only ethnolinguistic identification as Ukrainian (operationalized by the declaration that Ukrainian is one’s native language) is linked to perceptions of a thicker Russian-Ukrainian divide, with a full effect of about 1 point on the 11-point scale. Importantly, because we have controlled for nationality (civic identification with the Ukrainian state), individual language preference, and language embeddedness, we can be confident that it is in fact identification with a narrower ethnolinguistic vision of what it means to be Ukrainian that is associated with a stronger sense of “Russian Other,” not language preference, language use, or identification with a civic Ukrainian state. This makes intuitive sense and adds credence to our claim that these different indicators are
each capturing something substantially distinct.

[FIGURE 1 ABOUT HERE]

We now turn to the implications of these findings for how we reach conclusions about Ukrainian ethnicity’s effects on important attitudes, expectations, and behavior. With the exception of the overview provided in Table 3, for the sake of clear interpretation, we report—in what follows—the estimated effects of only the ethnic variables that are our primary focus, not of the other (non-ethnic) control variables. Nevertheless, it is important to keep in mind that these estimates are all reached while controlling for the same set of non-ethnic variables that are controlled for and reported in Figure 1: gender, age, education, urban residence, macroregion, and economic well-being. Full results can be found in the online appendix (Appendix Tables A3, A4, A5, and A6).

As a starting point, we report what someone would find if they arbitrarily included only one indicator of Ukrainian ethnicity in their survey-based studies of attitudes toward language policy and NATO, expectations of a Russian invasion, and participation in the EuroMaidan. Figure 2 graphically depicts the resulting estimated full effects of each measure on each dependent variable, with each column presenting results for one of our four dependent variables and each row giving results for one of our seven measures of Ukrainian ethnicity. Readers will readily see that as one moves from one dependent variable to the next, the effects of our seven indicators of ethnicity tend to vary together in ways suggesting they are all, to some degree, picking up the effects of an underlying notion of “Ukrainian ethnicity.” Thus if we look only at the point estimates (the dots) and not the 95% confidence intervals (the whiskers protruding from each dot), we find that virtually regardless of which single indicator is used to capture “Ukrainian ethnicity,” “Ukrainians” disproportionately tend to be against making Russian a second state language for the country as a whole, to support joining NATO, to expect a Russian invasion, and to be likely to join the Euromaidan protests.

[FIGURE 2 ABOUT HERE]

Beyond that, however, for all but one dependent variable, the specific choice of indicator can be decisive as to whether a given finding would qualify as substantively or statistically significant. First, let us look at substantive significance, or the magnitude of the estimated effects. Here we observe that different measures would lead us to draw differing conclusions about how powerful the effects of Ukrainian ethnicity are. For example, “Ukrainians” are about 25% less likely than others to back Russian as a state language according to the nationality scale, but this drops to just 15% if we use work language as our indicator instead.
Furthermore, and perhaps most consequentially, Ukrainian ethnicity would appear to be statistically insignificant as a predictor of many of these outcomes according to several of these indicators. Table 3 emphasizes this particular finding, summarizing the results regarding statistical significance of both the ethnic and control variables. All seven indicators are statistically significant only when it comes to opposing Russian as a second state language for Ukraine, the lone dependent variable in our study that is intrinsically related to ethnicity itself. When it comes to the other three dependent variables, only four of the seven are more than borderline statistically significant when it comes to backing NATO, only two are significant regarding joining the Euromaidan, and only one significantly predicts expecting a Russian invasion. We can clearly see that using one measure of ethnicity over another can greatly alter a study’s results. And even when the results are significant, we are still at a loss determining whether the detected association results from the substance of the measure itself or one of the other dimensions of ethnicity that are highly correlated with it.

[TABLE 3 ABOUT HERE]

To gain leverage on precisely what aspects of Ukrainian ethnicity are doing any causal work involved in these relationships, we now take our own advice and include a suite of four indicators: one measure of personal language preference (chosen survey language), one measure of language embeddedness (language of private life), ethnolinguistic identity (choice of native language), and self-declared nationality when forced to choose only one category. Figure 3 relates the findings (also see Table A7 in the online appendix). In the most general terms, we see that a personal language preference for Ukrainian, an ethnolinguistic identity as Ukrainian, and attachment to a civic Ukrainian nationality actually behave very similarly: all three are significant predictors (with roughly similar effect magnitudes) of opposing Russian as a second state language for Ukraine and backing NATO membership but are not significant in influencing expectations of a Russian invasion or participating in the Euromaidan protests.

[FIGURE 3 ABOUT HERE]

The measure that does not behave like the other three is the language used in private life, which is not a significant predictor of any attitude, expectation, or behavior reported in Figure 3. Since it must be interpreted here as the language used in private life controlling for one’s personal preference as to which language to use in communication, this variable should be interpreted as reflecting the effects of embeddedness in Ukrainian-speaking social environments. It appears, then, that the actual practice of communicating with close family members in Ukrainian—distinct from one’s own language preference—activates no considerations that are particularly important for the questions we are studying here. To reiterate, this
interpretation differs from how we must interpret results for the same variable in Figure 2, where (without other ethnic variables included as controls) it is to some degree capturing the effects of these other dimensions of ethnicity too and thus “behaving” in the statistical analysis more like the other measures.

What if we substitute language of private life with another indicator of language embeddedness, language of work? Figure 4 (and Table A8 in the online appendix) presents the answer. One might first observe that the point estimates of the size and direction of work language’s effects are almost exactly the same as those of private-life language on all four outcomes studied here, lending confidence that they are in fact accessing something similar. The primary difference is that the confidence intervals are narrower for the work language indicator, and this has the implication that language embeddedness now passes the statistical significance threshold as a predictor of two of our four dependent variables: joining the Euromaidan and opposing Russian as a second state language. This may well indicate that using a language at work creates more politically important forms of embeddedness in a language community than does using a language in private life, which many respondents may interpret to mean at home. Indeed, workplaces were important sites for the creation and bridging of the social network ties that Onuch and Sasse (2016) document were central to how individuals were mobilized to participate in Ukraine’s 2013–2014 protest wave, and it makes sense that Ukrainian-speaking workplaces would be more likely to mobilize than Russian-speaking ones. Whether a researcher should use private-life language or work language as their indicator of language embeddedness should depend on the specific topic under examination, but our study does offer some evidence that workplace embeddedness may be more politically meaningful.

[FIGURE 4 ABOUT HERE]

What happens if we keep Figure 3 as our baseline but use a different indicator of individual language preference? Figure 5 (see also Table A9 in the online appendix) reports what we find when we take the language someone is observed using in the survey as our indicator of individual language preference in place of the language chosen at the outset of the survey. This move has a surprising effect on attitudes toward language policy: individual language preference becomes insignificant in both instances where it had before been significant, as a predictor of opposition to Russian as a state language and support for joining NATO. A promising explanation is that the language someone initially chooses for the survey through the KIIS method is likely to be the purest indicator of the person’s actual language preference, the freest from social pressures. As the interview continues, interviewees are likely to pick up more cues from interviewers and adjust accordingly, blurring the measure.
Finally, we consider the implications of using a scale of Ukrainianness in place of the binary (forced-choice) measure of nationality reported in Figure 3. Figure 6 presents the results. The scale, it turns out, performs essentially like the binary measure with one major exception: the scale fails to be a statistically significant predictor of backing NATO membership. One interpretation is that such a scale is not so much capturing gradations of nationality as giving people an “easy middle” answer that they can use to escape reporting their true primary identification in the survey setting, meaning that the “between” responses reflect not true middle values on the nationality scale but a tendency not to take positions where middle options are available. The problem of such middle options has been widely documented in research on survey methodology, as noted earlier. Our tentative conclusion is thus that the nationality scale is a less useful indicator of true identification as Ukrainian than is the forced-choice measure.27

Conclusions

All told, we draw a number of important lessons from our findings, both methodological and substantive. First, we address implications for method. Perhaps most clearly, indicators of ethnicity should not be arbitrarily chosen when studying ethnicity’s effects. Even when different measures of ethnic identity are highly correlated, as they are in Ukraine, the inclusion of one over another can dramatically alter our interpretation of the magnitude of ethnicity’s effect on our dependent variable of interest or even whether a statistically significant effect is found at all.

Moreover, careful attention must be paid to control variables. Precise interpretation of ethnicity’s effects requires including control variables for each potentially salient dimension of identity that does not perfectly overlap with another; if a control variable is missing, one cannot tell which dimension of ethnicity is doing any causal work involved in a detected relationship between the ethnic variable and the dependent variable. For example, included without other ethnic controls, identifying Ukrainian as one’s native language significantly predicts joining the Euromaidan protests, but once we control for other dimensions of Ukrainian ethnicity, this relationship becomes insignificant and we are able to reveal instead that Ukrainian ethnicity’s influence on protest primarily involves the networked mobilization of Ukrainian-speaking workplaces (Figure 4).

For more precise interpretation of ethnicity’s effects in the Ukrainian context, we advise including four dimensions of ethnicity as either quantities of interest or
controls in one’s econometric models: individual language preference, language embeddedness, ethnolinguistic identity, and nationality. Even more precisely (but more tentatively), our findings suggest the following quantities be used, respectively, for these four dimensions: survey language chosen, language spoken at work, native language claimed, and forced-choice nationality. Since Ukraine’s ethnic context has been heavily shaped by the Soviet experience, especially when it comes to concepts such as “native language,” authors would most likely do well to apply this advice to studies of ethnicity’s political effects in other post-Soviet countries as well.

Our study also has a number of important substantive implications for how we understand ethnicity’s impact in Ukraine. First, we confirm the prior research finding that Ukraine is a divided society. But, second, we show that this is a highly blurry divide that is generally not accompanied by a sense of thick ethnic boundaries separating groups—even at a time of emerging interstate war pitting one group’s eponymous state (Ukraine) against the state identified most closely with largest other group (Russia). Third, the bluriness tends to break down into four dimensions: individual language preference, language embeddedness, ethnolinguistic identification, and nationality. Fourth, the dimension most associated with a thick boundary separating Russians from Ukrainians is ethnolinguistic identification. Fifth, attitudes toward language policy are robustly shaped by all four identity dimensions, though to somewhat different degrees. Sixth, support for NATO most strongly reflects language preference and self-identification with either the Ukrainian language or the Ukrainian state, but not language embeddedness. Seventh, expectations do not appear to be strongly patterned by ethnicity at all, at least when it comes to the perceived possibility of a Russian invasion as of May 2014. Finally, eighth, Euromaidan mobilization was “ethnic” primarily in the sense that it drew strongly from Ukrainian-speaking workplaces as nodal centers linking networks. Controlling for other factors, ethnolinguistic Ukrainians, Ukrainians by nationality, and people who preferred to speak Ukrainian were no more or less likely to join than were ethnolinguistic Russians, Russians by nationality, and people who preferred to speak Russian. We now urge other scholars to attempt to replicate these findings as well as to re-examine earlier conclusions reached without the full array of control variables that we recommend here both within Ukraine and beyond.

Acknowledgements

We would like to thank Gwendolyn Sasse, Grigore Pop-Eleches, Graeme Robertson, Volodymyr Kulyk, and Elise Giulano for their detailed suggestions on our paper at the ZOIS workshop on Identity in Ukraine, Berlin, Germany; to Joshua Tucker for his comments and feedback along with further comments provided by other participants at the 2017 ASN Convention; participants of the Princeton
Workshop on Ukraine 2017, and our collaborators in the composition of the UCEPS survey Timothy Colton and Nadiya Kravets. Olga Onuch would also like to thank HURI and Davis Center at the University of Harvard, which hosted her as a Research Fellow, thereby helping make the data collection and research possible. We are also grateful for the detailed feedback we received through the Post-Soviet Affairs review and editorial processes, with special thanks to editor Timothy Frye.

**Funding**

This work was supported by National Science Foundation grant SES-1445194.

**Notes**

1. Among many outstanding qualitative studies, see D’Anieri (2007), Hughes and Sasse (2002), Kuzio (2006), and Wilson (2002). Other outstanding works do employ quantitative analysis other than survey data in attempting to demonstrate the nature or salience of ethnic divides in Ukraine, using indicators such as patterns of rebellion against Soviet power or voting for parties with particular orientations (Darden forthcoming).

2. Focus groups can provide deeper insight into the meaning of identity dimensions, but always face a problem of generalizability.

3. Nationality (national’nist’; natsional’nost’) is sometimes translated in these studies as “ethnicity” (see discussions of this measure in Arel 2002 and Pirie 1996).

4. Some people, however, have been found to identify their “native language” as “the language of my country,” indicating a non-ethnic association, but this is not a dominant phenomenon (Kulyk 2018 [Query 3]).

5. On this danger regarding a related problem, see Esarey and Sumner (2017).

6. As long as sample sizes are adequate and correlations between any pair of these four variables are not perfect, multicollinearity is not a problem. Because each of these four indicators may be related to both the dependent and independent variables in distinct ways in the types of studies we are discussing here, to drop any one of these indicators is to introduce omitted variable bias (Epstein and King 2002, 204–207). If researchers do not have theoretical reason to believe any of these four dimensions of ethnicity lacks a distinct effect and if the number of observations turns out to be too small to distinguish among these indicators’ effects, forcing researchers to drop at least one of them, then authors should note that the data do not allow confident, precise interpretation of how exactly ethnicity is impacting outcomes and/or that all that can be established is a general role for
ethnic factors (leaving the mechanism unclear).

7. Of course, this is no analytical panacea. As Kulyk’s (2018) article in this special issue makes clear, the precise meaning of different ethnic markers as personal points of reference can themselves change over time with a changing environment. To the extent this is true, results even with our recommended controls need to be interpreted accordingly.

8. Minus the already-annexed Crimea but including Donetsk and Luhansk, where insurgent forces had not yet consolidated control.

9. In our frequency calculations and regression analyses, we utilize probability weights calculated by KIIS to bring each wave in line with official 2013 population statistics for age, gender, and region.

10. After each interview, interviewers were asked to report: What language did the respondent answer you in? (1) Ukrainian; (2) In most cases Ukrainian, but sometimes in Russian; (3) Half one language and half the other; (4) A mix of Ukrainian-Russian language (surzhyk); (5) In most cases, Russian, but sometimes Ukrainian; (6) In Russian.

11. “Now let us talk a little bit about language. Please tell me … (1) Which language do you typically speak in your private life? If you speak several languages in your private life, please, tell me, which one you consider the main one. (2) Which language do you typically speak at work: Ukrainian, Russian, Other, Ukrainian and Russian equally?”

12. “Now let us talk a little bit about language. Please tell me … (3) What language do you consider your native language: Ukrainian, Russian, Other, Ukrainian and Russian equally?’

13. “Some people belong only to one nationality, while others consider themselves to be of mixed nationality. Please tell me, to approximately what degree to you belong to the following groups? (1) Russian (2) Ukrainian (3) other.” Response options were: entirely; mostly; half; partly but less than half; none.

14. “If you had to register only one nationality, which would you choose? Russian, Ukrainian, Other (please specify: _________________)’

15. “Please tell me to what extent you agree or disagree with the following statements … The Russian language should be given the status of a second state language in Ukraine, equal in status to Ukrainian.” Responses included completely agree, tend to agree, tend to disagree, and completely disagree. We code this as those who agree and tend to agree (1) versus all others (0).
16. “Due to the deterioration of relations between Ukraine and Russia, many argue that Ukraine should strengthen its security. Please tell me if you agree or disagree with the following statements: ... Ukraine should join NATO.” Responses included completely agree, tend to agree, tend to disagree, and completely disagree. We code this as those who agree and tend to agree (1) versus all others (0).

17. “Please tell me to what extent you agree or disagree with the following statements ... It is likely that in the near future, Russian troops will try to take some parts of Eastern Ukraine.” Responses included completely agree, tend to agree, tend to disagree, and completely disagree. We code this as those who agree and tend to agree (1) versus all others (0).

18. “Since autumn of last year, how did you participate in any of the following demonstrations?” Among the items were “Euromaidan in Kyiv” and “EuroMaidan in your native region.” Possible answers included: never, once, and more than once. We coded all respondents who said they had participated in this particular protest movement once or more than once as 1 and all others 0.

19. We utilize a six-point scale for education: none or elementary; incomplete secondary; secondary; secondary specialized; incomplete higher; and higher.


21. “In general, did your family win or lose as a result of the economic changes that have taken place since Ukraine became an independent country?” Responses included: won; mostly won; mostly lost; and lost. Volunteered responses of “won some and lost some” were also coded. Those answering that they “won” or “mostly won” were coded as transition winners.

22. We utilize a seven-point scale for responses to the question: “Which of the following statements best describes the financial situation of your family?”: “We do not have enough money even for food”; “We have enough money but only for the most necessary things”; “We have enough money for daily expenses, but to even buy clothes is difficult”; “Usually, we have enough money, but to buy expensive things, such as, for example, a refrigerator, a TV and a washing machine, it takes a longer time, we have to borrow or get credit”; “We can afford expensive purchases
without too much difficulty, but buying a car is still beyond our means”; “We can buy a car without much effort, but buying a home is still difficult”; “At the present time we can afford anything we want.”

23. We rescale all non-binary variables to a range of 0 to 1, with 0 and 1 representing the minimum and maximum values, respectively, of each variable observed.

24. This suggests that the line between individual language preference and actual language practice can become fuzzy, especially in realms where the individual has more opportunity to choose or influence one’s interlocutors, as in private life, home, and family.

25. Results in tabular form can be found in Table A2 of the online appendix.

26. It also bears noting here that this “othering” is a strongly western Ukrainian phenomenon, against which baseline all other macroregions of Ukraine (especially eastern Ukraine) stand out for seeing more similarity—even controlling for all four ethnic variables in our study. People with deeper pre-independence personal histories (that is, older people) also stand out for seeing thinner ethnic boundaries, while those who have benefitted materially from the independence period are more likely to perceive thicker distinctions—an interesting economic effect that we leave for future work to interpret.

27. That said, we do not rule out certain other ways of conceptualizing nationality in Ukraine that we leave for other research. First, perhaps people can consider themselves both fully Ukrainian and fully Russian, an option not given in our survey but documented, for example, regarding Spanish and Catalan (Stepan, Linz, and Yadav 2011). Second, perhaps people who combine Russian with Ukrainian identity display behavior that is not simply a linearly weaker version of the expected Russian/Ukrainian behavior, but behavior that is completely different somehow.

References


**Tables**

*Capturing Ethnicity: The Case of Ukraine*

*Prepared for Post Soviet Affairs, special issue on ‘Identity Politics, Crisis, and Conflict: The Critical Case of Ukraine’*
### Table 1. Key Studies of Ethnicity’s Effects in Ukraine Including Multiple Measures

<table>
<thead>
<tr>
<th>Authors</th>
<th>(Barrington 2002)</th>
<th>(Barrington and Faranda 2009)</th>
<th>(Kulyk 2011)</th>
<th>(Shulman 2005b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variables</strong></td>
<td>Support for government; regime</td>
<td>Attachment to Russia</td>
<td>Attitudes to language policy; foreign policy; Bandera</td>
<td>Attitudes to economic &amp; political systems</td>
</tr>
<tr>
<td><strong>Nationality measure</strong></td>
<td>Forced choice of Ukrainian, Russian, or Other “nationality”</td>
<td>Forced choice of Ukrainian, Russian, or Other “nationality”</td>
<td>Forced choice of Ukrainian or Russian “nationality” (apparently)</td>
<td>Forced choice of Ukrainian, Russian, or Other “nationality”</td>
</tr>
<tr>
<td><strong>Language measure</strong></td>
<td>Interviewers’ observation of respondent’s language patterns in answering survey questions</td>
<td>Language spoken at home</td>
<td>Main language of everyday use; and Native language</td>
<td>Language spoken at home</td>
</tr>
<tr>
<td><strong>Main findings</strong></td>
<td>Significance of language and nationality varies by dependent variable and dissipates with regional controls</td>
<td>Language is significant; significance of nationality varies by region</td>
<td>Nationality and native language are significant; language is not</td>
<td>Nationality and language insignificant or weak</td>
</tr>
<tr>
<td></td>
<td>Language chosen for survey</td>
<td>Language observed in survey</td>
<td>Language in private life</td>
<td>Language at work</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>-------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Language chosen for survey</td>
<td>1</td>
<td>0.760***</td>
<td>0.690***</td>
<td>0.508***</td>
</tr>
<tr>
<td>Language observed in survey</td>
<td>0.760***</td>
<td>1</td>
<td>0.803***</td>
<td>0.607***</td>
</tr>
<tr>
<td>Language in private life</td>
<td>0.690***</td>
<td>0.803***</td>
<td>1</td>
<td>0.661***</td>
</tr>
<tr>
<td>Language at work</td>
<td>0.508***</td>
<td>0.607***</td>
<td>0.661***</td>
<td>1</td>
</tr>
<tr>
<td>Native language</td>
<td>0.559***</td>
<td>0.628***</td>
<td>0.701***</td>
<td>0.506***</td>
</tr>
<tr>
<td>Nationality binary</td>
<td>0.269***</td>
<td>0.279***</td>
<td>0.318***</td>
<td>0.222***</td>
</tr>
<tr>
<td>Nationality scale</td>
<td>0.282***</td>
<td>0.298***</td>
<td>0.334***</td>
<td>0.233***</td>
</tr>
</tbody>
</table>

*** p < 0.001
Table 3. Summary Of Findings (Significance, Direction Of Effect) From 28 Logit Regression Analyses, 7 For Each Of 4 Dependent Variables, With Each Model Containing The Standard Set Of Control Variables With One Of Seven Ethnic Independent Variables

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Support Russian As Official State Language</th>
<th>Support NATO Joining</th>
<th>Expecting Russian Invasion</th>
<th>Participating In Euromaidan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ukrainian Language Survey Chosen</td>
<td>YES -</td>
<td>YES +</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Ukrainian Language Observed in Survey</td>
<td>YES -</td>
<td>YES +</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Ukrainian Language in Private Life</td>
<td>YES -</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Ukrainian Language at Work</td>
<td>YES -</td>
<td>NO</td>
<td>YES +</td>
<td>YES +</td>
</tr>
<tr>
<td>Ukrainian Native Language</td>
<td>YES -</td>
<td>YES +</td>
<td>NO</td>
<td>YES +</td>
</tr>
<tr>
<td>Ukrainian Nationality</td>
<td>YES -</td>
<td>YES +</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>East</td>
<td>YES +</td>
<td>YES -</td>
<td>YES -</td>
<td>YES -</td>
</tr>
<tr>
<td>South</td>
<td>YES +</td>
<td>YES -</td>
<td>YES -</td>
<td>YES -</td>
</tr>
<tr>
<td>Center</td>
<td>DEPENDS*</td>
<td>YES -</td>
<td>YES -</td>
<td>YES -</td>
</tr>
<tr>
<td>Transition Winner</td>
<td>YES -</td>
<td>YES +</td>
<td>NO</td>
<td>YES +</td>
</tr>
<tr>
<td>Female</td>
<td>NO</td>
<td>YES -</td>
<td>YES -</td>
<td>NO</td>
</tr>
<tr>
<td>Family Financial Situation</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES +</td>
</tr>
<tr>
<td>Education</td>
<td>NO</td>
<td>YES +</td>
<td>NO</td>
<td>YES +</td>
</tr>
<tr>
<td>Age</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES -</td>
</tr>
<tr>
<td>Urban Locality</td>
<td>NO</td>
<td>YES -</td>
<td>NO</td>
<td>YES +</td>
</tr>
</tbody>
</table>

Yes = significant at the 95% level (p<.05)  
No= Not significant at the 95% level (p≥.05)  
+ = Positive relationship  
- = Negative relationship
*Significance varies depending on which ethnicity variable is included. The result is YES + when either forced-choice nationality or native language is the ethnic variable in the equation, NO for all other ethnic variables.

### List of Figures

**Figure captions**

Figure 1. Full effect of factors on 11-point scale of perceived ethnic boundary thickness between Ukrainians and Russians.
Figure 2. Full effects of ethnicity on probability (%) of...
Figure 3. Full effects of four ethnicity variables on probability (%) of … language of private life = indicator of language embeddedness.
Figure 4. Full effects of four ethnicity variables on probability (%) of … language of work = indicator of language embeddedness.
Figure 5. Full effects of four ethnicity variables on probability (%) of … language observed = indicator of language embeddedness.
Figure 6. Full effects of four ethnicity variables on probability (%) of ... scale of Ukrainianess = measure of nationality.