War and Nation-Building: Evidence from Ukraine*

Serhii Abramenko† Vasily Korovkin‡ Alexey Makarin§ Moses Shayo¶

February 2026

Abstract

How quickly and how deeply can violent conflict reshape national identity? We study this question in the context of Ukraine and the Russia–Ukraine war, using an unusually comprehensive set of data that spans ethnic, linguistic, and cultural dimensions of identity. We document large and persistent shifts toward Ukrainian identity following both major conflict episodes, away from Russian and mixed identities. In a series of nationally representative surveys, Ukrainian ethnic identification rose by about 10 percentage points after the 2014 conflict and by a further 7 percentage points after Russia's full-scale invasion in 2022. Consistent with national consolidation rather than polarization, the largest shifts occurred in districts that were historically more ethnolinguistically Russian. Language use evolved in parallel: changes after 2014 were modest, but in 2022 roughly 20 percent of respondents switched to Ukrainian in survey interviews, a pattern confirmed using within-user evidence from Twitter and private Google searches. Cultural consumption and intergenerational choices also shifted sharply after 2022, toward Ukrainian cuisine, Ukrainian-language music, and traditionally more Ukrainian baby names. These changes persist after accounting for compositional change and population displacement. Together, the results show that national identity is far more malleable than commonly assumed: large-scale violence can generate rapid, behaviorally meaningful, and broadly shared identity change—even among populations with substantial pre-war cultural and linguistic ties to the aggressor.

JEL: D72, D74, O57, P00, Z13

Keywords: nation building, identity, violent conflict, war, cultural change, Ukraine

*We thank Daron Acemoglu, Ruben Enikolopov, Ben Olken, Panu Poutvaara, Nancy Qian, Tavneet Suri, Nico Voigtlander, Romain Wacziarg, Ekaterina Zhuravskaya, and conference participants at the AEA Annual Meeting, the ifo Workshop on Reconstruction after Conflict, and the SIOE Annual Conference for helpful comments. We thank Andrey Samarskiy and Aleksander Vedishchev for superb research assistance, and Valerio Sergio Castaldo for contributions at earlier stages of the project.

†Rome Economic Doctorate (RED) (e-mail: sabramenko@redphd.it).

‡Universitat Pompeu Fabra and CEPR (e-mail: vasily.korovkin@upf.edu).

§MIT Sloan School of Management, CEPR, and CESifo (e-mail: makarin@mit.edu).

¶The Hebrew University of Jerusalem and CEPR (e-mail: mshayo@huji.ac.il).

"A common enemy is the best catalyst for forging a common identity."

— Yuval Noah Harari, 21 Lessons for the 21st Century

"Give people a common enemy, and you will give them a common identity. Deprive them of an enemy and you will deprive them of the crutch by which they know who they are."

— James Alison

Introduction

When Russia annexed Crimea in March 2014, both the Russian government and many Western analysts predicted that Ukraine's population—with nearly a third of the country reporting Russian as a native language—would fracture along ethnolinguistic lines. Russian military and political strategy at the time was explicitly premised on such fragmentation, particularly in the country's eastern and southern regions. This paper asks whether these predictions were borne out. A decade later, the opposite pattern has emerged. Rather than fragmenting, Ukrainian national identity has consolidated, with the largest shifts occurring precisely in regions where ethnic Russians were most concentrated before the war. This consolidation intensified following Russia's full-scale invasion in 2022, generating especially large and persistent shifts across ethnic identification, language use, and cultural behavior.

This outcome matters beyond the Ukrainian case. National cohesion is a key determinant of public goods provision, the capacity to sustain collective defense, and long-run economic development (Rohner and Zhuravskaya, 2023). Yet whether armed conflict ultimately consolidates or fragments national identity remains theoretically ambiguous. External conflict can strengthen national identification by increasing the salience and perceived status of the nation, fostering in-group cooperation, collective mobilization, and "rally-around-the-flag" effects (Mueller, 1970; Bauer et al., 2016).

Perceived military success or international recognition may further reinforce national identity (Sambanis et al., 2015). At the same time, conflict can weaken national identification by exacerbating internal divisions when violence overlaps with pre-existing ethnic or linguistic cleavages. In such cases, heightened perceptions of intra- and inter-group differences—especially when internal minorities are culturally or linguistically close to the external adversary—may activate subnational identities and generate polarization (Shayo and Zussman, 2011; Sambanis and Shayo, 2013). How these competing forces interact in externally threatened but internally heterogeneous societies therefore remains an open empirical question.

Ukraine provides an unusually sharp setting for testing these opposing predictions. Prior to 2014, a substantial share of the population was ethnolinguistically close to the aggressor state: in 2012, roughly 30 percent of Ukrainians reported some Russian ethnic identity, and a majority used Russian or a Russian–Ukrainian mix in daily communication. Political cleavages were also pronounced, with deeply divided public opinion on European integration and a significant presence of pro-Russian parties in national politics as late as 2021. Russian cultural and information products were widely consumed, further blurring national boundaries. Taken together, these features made both consolidation and polarization plausible ex ante. This ambiguity was not merely academic: Russian political and military strategy was premised on the expectation of local support in Russian-speaking regions, while many Western governments underestimated the scale of Ukrainian resistance (Boulègue et al., 2023).

This paper studies how two discrete conflict shocks—the onset of hostilities in 2014 and Russia's full-scale invasion in 2022—reshaped national identity in Ukraine. Our empirical strategy combines two complementary approaches. First, we exploit the large scale and unexpected nature of both shocks in a pre–post design to test whether national identity exhibits sharp breaks at conflict onset. Second, we implement a difference-in-differences design that leverages continuous variation in pre-war ethnolinguistic composition across districts, measured by the share of ethnic Russians

in the 2001 Census.[1] This second approach directly distinguishes consolidation from polarization by testing whether identity shifts following conflict onset are larger or smaller in areas with stronger historical ties to Russia.

Across all outcomes, we find clear evidence of identity *consolidation* rather than *polarization*. Both conflict episodes generate sharp shifts toward Ukrainian national identity, and these shifts are systematically larger in regions where pre-war attachment to Ukrainian identity was weakest.

We begin by examining changes in ethnic self-identification. Following the onset of hostilities in 2014, Ukrainian ethnic identification increases sharply by approximately 10 percentage points, accompanied by a corresponding decline in mixed and Russian identities. After Russia's full-scale invasion in 2022, Ukrainian identification increased by a further 7 percentage points, reaching 95 percent and leaving fewer than 2 percent of respondents reporting a mixed or Russian identity. Importantly, these shifts are largest in districts with higher pre-war shares of ethnic Russians, directly contradicting predictions of polarization. Event-study estimates show sharp breaks at each conflict onset and no differential pre-trends across districts, supporting the identifying assumption underlying our difference-in-differences design.

Language behavior exhibits a closely parallel but more pronounced pattern. Changes after 2014 are modest, but the 2022 full-scale invasion triggers a large and abrupt shift toward Ukrainian language use. In survey interviews, the share of respondents answering in Ukrainian rises by more than 20 percentage points following the invasion. As with ethnic identification, these shifts are largest in regions with higher pre-war concentrations of ethnic Russians. Evidence from additional data sources corroborates this pattern and helps rule out compositional change or social desirability as primary explanations. Using the universe of original tweets from Ukraine, we document sharp increases in Ukrainian language use even when controlling for user fixed effects, thus isolating within-individual changes. Google search data show similar

[1] Results are unchanged if we instead use the share of Russian native speakers by district in the 2001 Census, the most recent census available for Ukraine.

shifts toward Ukrainian-language queries, indicating that language change extends to private settings unlikely to be driven solely by interviewer effects or social desirability.

Cultural consumption provides particularly sharp evidence of identity change. Using Spotify's weekly Top-200 charts for Ukraine, we document a large reorientation of music consumption toward Ukrainian-language content following the 2022 invasion: the share of Ukrainian-language songs rises from about 5 percent before the invasion to nearly 30 percent afterward. This shift reflects changes in both listener demand and artist behavior and cannot be explained by compositional change or population displacement. Beyond music, we observe parallel reorientation in other private and persistent behaviors. Searches increasingly favor Ukrainian over Russian cuisine, and parents become significantly more likely to choose Ukrainian rather than Russian baby names. Together, these patterns show that conflict-induced identity change extends beyond stated preferences and public-facing behavior to everyday private consumption choices and intergenerational decisions.

Taken together, the results point to a broad-based and persistent shift toward Ukrainian national identity. More broadly, they show that national identity is far more malleable than commonly assumed: large-scale external violence can generate rapid, behaviorally meaningful, and widely shared identity change, even in historically heterogeneous societies with pre-existing cultural and linguistic ties to the adversary.

This paper contributes to the literature on conflict and nation-building by providing a clean test of whether external war consolidates or polarizes national identity in heterogeneous societies. While existing work documents heterogeneous effects of conflict on political cohesion and social outcomes (Dell and Querubin, 2018; Dehdari and Gehring, 2022; Ananyev and Poyker, 2023), it rarely isolates consolidation and polarization as competing mechanisms within the same conflict, particularly in settings where both outcomes are plausible ex ante. Exploiting continuous variation in pre-war ethnolinguistic composition, we show that identity shifts following external conflict are strongest precisely where pre-war attachment to the nation was weakest. We further show that these identity shifts translate into large changes in consumption behavior,

linking conflict-induced identity change to economically meaningful outcomes.

The paper also contributes to the literature on national identity and nation-building by expanding both the scope and measurement of identity change. A growing body of work studies how policies, institutions, and historical shocks shape national identification and language use (Jha, 2013; Aghion et al., 2019; Bazzi et al., 2019; Alesina et al., 2021; Assouad, 2025; Bazzi et al., 2025; Carlitz et al., 2025; see Rohner and Zhuravskaya, 2025 for a recent comprehensive overview), while theory emphasizes endogenous identity choice driven by similarity, status, and strategic incentives (Shayo, 2009; Sambanis and Shayo, 2013). Much of this literature relies on survey-based measures, leaving open concerns about social desirability and behavioral relevance. We address these concerns by combining surveys with private-action outcomes—music consumption, search behavior, and naming choices—that are costly, persistent, and less susceptible to reporting bias. Methodologically, our approach illustrates how integrating high-frequency surveys with large-scale digital traces can substantially strengthen inference about identity formation.[2]

The remainder of the paper is organized as follows. Section 1 describes the Ukrainian context. Section 2 presents the data sources. Section 3 reports the empirical results for ethnic and linguistic identity. Section 4 examines cultural identity outcomes. Section 5 concludes.

1 Ukrainian Context

1.1 Ethnic and Linguistic Identity in Ukraine

In the Ukrainian context, the Soviet concept of "nationality" is commonly used in place of ethnicity, as reflected in the 2001 census and in the surveys we analyze. Under

[2]Related work on the Russia–Ukraine war studies economic disruptions (Korovkin and Makarin, 2023; Korovkin et al., 2025), refugee flows (Harm Adema et al., 2024), and documents shifts in identity and language use following earlier phases of the conflict (Kulyk, 2016, 2018, 2024; Abramenko et al., 2024). Büyükeren (2025) shows that Ukraine's decentralization reforms boosted Ukrainian identity in rural areas. We complement this literature by providing unified evidence across both major conflict shocks, by directly testing consolidation versus polarization using granular within-country variation in pre-war ethnolinguistic composition made possible by newly accessed city-level KIIS data, and by showing that identity shifts extend beyond stated attitudes to private behavior and long-term choices.

Soviet rule, nationality was determined by parental origin and recorded as a mandatory attribute in civil passports. This legacy renders contemporary self-reported nationality in Ukraine conceptually broader than ethnicity alone: changes may reflect shifts in ethnic self-identification or in the relative salience of in the perceived importance of ethnicity relative to citizenship (Barrington, 2018; Kulyk, 2018). Throughout the paper, we therefore use the terms ethnicity and nationality interchangeably, while recognizing this conceptual overlap. Because Ukrainian citizenship is largely invariant for individuals in our sample, shifts in self-reported nationality reflect changes in self-identified ethnicity and the perceived relevance of ethnicity vis-à-vis the civic component—changes that are observationally equivalent.

High rates of ethnic intermarriage in Ukraine (25% in 1989; Rapawy, 1997), particularly in the east and south, further contribute to the fluidity of ethnic identification, making Ukraine a fertile setting for identity change (Jia and Persson, 2021).

Language complicates matters further. During the late 19th and 20th centuries, Russification shaped education, media, and public life (Kravtsiv and Kubijovyc, 1993). The Valuev Circular (1863) banned most Ukrainian-language publications; the Ems Ukaz (1876) extended these restrictions to schools and theater (Subtelny, 2009, 282–84). Yet Ukrainian remained the predominant spoken language in rural areas and retained its vitality as a marker of local identity (Magocsi, 1996, 479–480). After 1917, early Soviet Ukrainization briefly reversed the trend: by 1929, over 80% of schools taught in Ukrainian (Subtelny, 2009, 389–90). Stalin's reversal in 1933 ended this interlude (Subtelny, 2009, 422).

The subsequent decades deepened Ukraine's linguistic Russification. The Holodomor (1932–33) and World War II sharply reduced Ukraine's population, and large-scale resettlement of Russians into the republic transformed its ethnic balance (Markevich et al., 2025; Subtelny, 2009, 415–16, 480–85). By the Brezhnev era, Russian language had become the key to mobility and higher education, while Ukrainian publishing shrank: the share of books in Ukrainian fell from 60% to 24% from 1958 through 1980

(Subtelny, 2009, 536).[3]

As a result, numerous ethnic Ukrainians adopted Russian as their primary language. The share of Ukrainians in the USSR who reported Ukrainian as their native tongue fell from 87.7% in 1959 to 81.1% in 1989. Even after independence, Russian remained prevalent in media, education, and daily communication. According to the 2001 Census, 87.8% of the population were fluent in Ukrainian and 67.7% in Russian.[4] The onset of the conflict made Ukrainians, especially Russian speakers, increasingly perceive linguistic similarities with Russia as a security liability, since Moscow invoked the need to "protect" Russian speakers to justify aggression (Monde, 2024).

In Ukraine, nationality and language are less tightly aligned than in most European nation-states (see, e.g., Kulyk and Hale, 2022). It is therefore useful to distinguish three broad ethnolinguistic groups—Ukrainian-speaking Ukrainians, Russian-speaking Ukrainians, and predominantly Russian-speaking ethnic Russians—as proposed in Arel and Khmelko (1996). *Ex ante*, changes in national and linguistic identity should be concentrated among the latter two groups, while Ukrainian-speaking Ukrainians have more entrenched identities and thus less scope to adjust in this context.

Two clarifications are important for interpreting identity responses to the war. First, conflict plausibly permits identity change in multiple directions. On one hand, prior work shows that the 2014 onset of hostilities strengthened attachment to Ukraine and reduced Russian language use (Pop-Eleches and Robertson, 2018; Abramenko et al., 2024; Kulyk, 2024). On the other hand, Russian political narratives explicitly denied the existence of a distinct Ukrainian nation and sought to activate subnational identities among Russian-speaking Ukrainians, particularly in the east and south, through appeals to shared origins and threats of linguistic repression (Bækken, 2024). Second, ethnolinguistic identity and language use vary sharply across space: Russian nationality and language are more prevalent closer to Russia and the Black Sea, while

[3]Another dimension of this linguistic landscape is *Surzhyk*, a mixed Ukrainian–Russian vernacular long stigmatized as "incorrect" speech. Since the 2022 invasion, Surzhyk has seen a revival, with many Ukrainians embracing it as an inclusive marker of identity (Osypenko, 2025).

[4]State Statistics Committee of Ukraine, All-Ukrainian Population Census 2001, http://2001.ukrcensus.gov.ua/eng/results/general/language/.

Ukrainian identification and language use increase along an east–west gradient. This spatial structure implies that regional patterns of change in nationality and language are informative about which ethnolinguistic groups drive aggregate identity shifts, a feature we exploit in our empirical analysis.

1.2 A Brief Account of the Russia–Ukraine War (2014–24)

The Outbreak (2014–15). Mass protests began in Ukraine in November 2013 after President Yanukovych declined to sign the EU–Ukraine Association Agreement. By February 2014, the protests culminated in a change of government, with President Yanukovych leaving the country and the Verkhovna Rada assuming executive authority. In March 2014, Russia annexed Crimea following the deployment of armed forces and a local referendum that was not internationally recognized. Shortly thereafter, armed groups seized administrative buildings in parts of the Donetsk and Luhansk regions and proclaimed the establishment of "people's republics." Although the Russian government denied direct involvement, multiple investigations documented the presence of Russian nationals and regular military units in key engagements, including Ilovaisk and Debaltseve. Ukraine mobilized its armed forces but did not reestablish control over these territories (Plokhy, 2015, chap. 27; D'Anieri, 2023, chap. 7).

A Frozen Conflict (2015–22). In February 2015, Ukraine, Russia, and representatives of the self-proclaimed separatist authorities signed the Minsk II agreement, which established a ceasefire but left core political and territorial issues unresolved. Following the agreement, Russian-backed separatists retained control over roughly one-third of the Donetsk and Luhansk oblasts, including their regional capitals. While large-scale fighting subsided, low-intensity violence—including shelling, sniper fire, and periodic escalations—continued throughout the period, resulting in ongoing military and civilian casualties, albeit at substantially lower rates than in 2014–15. Implementation of the Minsk framework stalled, and the frontline remained largely unchanged, producing a prolonged frozen conflict (Plokhy, 2015, chap. 27; D'Anieri, 2023, chap. 8).

The Full-Scale Invasion (2022–24). On 24 February 2022, Russia launched a full-scale invasion of Ukraine. Military operations initially targeted Kyiv and other major cities, but by April 2022 Russian forces withdrew from northern Ukraine following sustained Ukrainian resistance. The war subsequently shifted to the east and south, where Russian forces occupied additional territory in Donetsk and Luhansk oblasts and advanced into Zaporizhzhia and Kherson oblasts. In late 2022, Russia formally claimed the entirety of all four oblasts as its territory, despite not exercising full control over them. Ukrainian counteroffensives later that year recaptured most of Kharkiv oblast and the city of Kherson, while subsequent operations in 2023 produced more limited territorial changes. By 2024, the conflict had largely transitioned into a war of attrition characterized by sustained fighting along established lines, with Russia making gradual territorial advances (D'Anieri, 2023, chap. 9; CFR, 2024).

Consequences of the War. The conflict has had large economic and demographic consequences. The 2014 hostilities led to a substantial contraction of economic activity, on the order of 20–50 percent, in separatist-controlled areas (Kochnev, 2019), as well as additional losses of about 6–7 percent in non-conflict areas due to the disruption of domestic production networks and Russia–Ukraine trade (Korovkin and Makarin, 2023; Korovkin et al., 2025). Following the 2022 full-scale invasion, Ukraine's GDP declined by approximately 31 percent (Federle et al., 2024). The war has also resulted in severe human losses: according to the Center for Strategic and International Studies, by late 2025, cumulative casualties were estimated at roughly 400,000 Ukrainians and one million Russians (Mayer et al., 2026). Population displacement has been extensive, with about 3.6 million Ukrainians internally displaced and an additional 6.8 million seeking refuge abroad, primarily in the European Union. In parallel, military strikes have damaged critical infrastructure, including energy systems and transport networks (CFR, 2024; UN High Commissioner for Refugees, 2025).

2 Data

We combine repeated quarterly surveys of the Ukrainian population with administrative data and digital records of language use and consumption. The surveys provide direct measures of ethnic identification and language choice, while the other datasets capture revealed behavior in private or high-stakes settings that are less susceptible to reporting or interviewer effects.

2.1 Survey Data: KIIS Omnibus Surveys

Our primary data source is the omnibus survey conducted by the Kyiv International Institute of Sociology (KIIS). We use 72 survey waves fielded quarterly between 2008 and 2024, comprising more than 156,000 respondents. Each wave is designed to be nationally representative of the adult population residing in government-controlled territories at the time of the survey.

The surveys contain rich information on respondents' demographic characteristics, place of residence, ethnic self-identification ("nationality"), and language use. Crucially for our analysis, we obtained access to restricted KIIS records that identify respondents' city or settlement of residence, allowing us to link individuals to pre-war local ethnolinguistic composition at a fine geographic level. This feature enables within-country heterogeneity analysis that is not possible using the coarser regional identifiers available in publicly released KIIS data.

We focus on three survey-based identity outcomes that are consistently available over time. First, respondents answer an open-ended question on self-reported nationality. Second, for respondents identifying as Ukrainian and/or Russian, the survey includes an ordered five-point nationality scale (available through early 2022) that captures mixed and intermediate identities. Third, interviewers record the language of the interview (Ukrainian, Russian, or mixed) following a standardized protocol in which respondents choose the language in which they feel most comfortable. This measure reflects revealed linguistic choice in a low-stakes interaction rather than a direct attitudinal response.

Further details on the data and the exact wording of the ethnic and linguistic identity questions are provided in Appendix B.1.

2.2 Twitter Data: Public Language Use

To measure public linguistic behavior at high frequency and within individuals, we use the universe of original (non-retweeted) tweets geolocated to Ukraine between January 2012 and November 2022, comprising approximately 36 million tweets. Language is identified using Twitter's built-in language classifier, which performs comparably to standard open-source tools. While geotagging became less common over time, the number of precisely located tweets remained broadly representative in terms of user composition and linguistic behavior. See Abramenko et al. (2024) for further details.

We exploit user identifiers to estimate specifications with account fixed effects, so identification comes from within-user changes in language choice over time. This design rules out changes driven by shifts in the composition of users and provides a behavioral validation of the survey evidence in a public communication setting.

2.3 Google Search Data: Private Language Use and Consumption

We complement survey and social media evidence with Google Trends data, which capture private information-seeking behavior. Because search activity is typically private and not interviewer-mediated, these data are less susceptible to social desirability or strategic reporting concerns. In 2021, Google was overwhelmingly the most popular search engine in Ukraine, accounting for over 88% of total search engine market share across all platforms.[5]

We use Google Trends to measure two dimensions of identity-related behavior. First, we construct quarterly region-level measures of the relative use of Ukrainian versus Russian in search queries from 2012 to 2025. These measures are based on pairs of Ukrainian and Russian spellings of the same words and capture revealed language choice in private settings (see Appendix B.2.1 for details).

[5]StatCounter Global Stats, "Search Engine Market Share in Ukraine (2021)," available at https://gs.statcounter.com/search-engine-market-share/all/ukraine/2021.

Second, we use search activity to study cultural consumption, focusing on interest in national cuisine. We assemble an oblast-by-month panel of search intensity for major Ukrainian and Russian dishes from 2012 to 2024 and examine whether interest shifts differentially toward dishes more strongly associated with Ukrainian culture following the war (see Appendix B.2.2 for details).

In both applications, identification relies on within-region changes over time and exploits sharp breaks at the onset of hostilities in 2014 and the 2022 invasion.

2.4 Baby Names: Intergenerational Identity

To capture long-term and intergenerational identity choices, we use administrative data from the Ukrainian Ministry of Justice on popular baby names by region and half-year from 2015 onward. Each name is assigned a time-invariant "Ukrainianness" score based on its relative pre-war popularity in Ukraine versus Russia. We study how the popularity of more Ukrainian versus more Russian names changes after the 2022 invasion, controlling for region-time fixed effects.

2.5 Music Consumption: Spotify Charts

Finally, we analyze cultural consumption through the lens of popular music. Our empirical analysis draws on four complementary data sources. First, we use Spotify Weekly Top-200 charts for Ukraine (July 2020–August 2025), scraped from Kworb, which provide weekly chart positions and streaming counts for all ranked tracks. Second, we collect song-level metadata — including release dates and artist information — from the Spotify Web API. Third, we retrieve song lyrics from Genius using the Genius API , which allows us to classify tracks by language using automated language detection. Fourth, for the supply-side analysis, we construct a comprehensive catalogue of Ukrainian artists by combining curated lists from NUAM Club, Spotify, and Wikipedia contributors, and then retrieve their full discographies from Genius.

The final dataset contains 3,850 distinct charting tracks performed by 572 artists, together with song-level information on language, release date, chart trajectory, and total streams. This structure allows us to document the overall shift in music consump-

tion, isolate demand-side changes by focusing on the chart performance of pre-war releases, and identify supply-side adjustments in language choice among Ukrainian artists.

2.6 Pre-War Ethnolinguistic Composition

To measure pre-existing cultural proximity to Russia, we use district-level ethnolinguistic composition from the 2001 Ukrainian Census, the most recent census available. Our main measure is the share of residents reporting Russian nationality in each pre-2020 district (rayon). In robustness checks, we alternatively use the share of Russian native speakers, with similar results.

Ukraine's administrative reform in 2020 consolidated districts into a smaller number of larger units. To maintain consistent treatment intensity, we map post-2020 survey respondents back to their corresponding pre-reform districts and use the 2001 composition measure throughout. This allows us to combine pre-war ethnic composition with contemporaneous district fixed effects and clustering.

2.7 Population Displacement and Local Controls

To address concerns about compositional change due to migration and displacement, we incorporate data on internally displaced persons (IDPs). We use oblast-level IDP statistics from the Ukrainian Ministry of Social Policy for 2014–2019 and district-level data for 2020 onward, adjusting for administrative changes. For the period following the 2022 full-scale invasion, we rely on district-level IDP dynamics from the International Organization for Migration (IOM). These measures allow us to control for local displacement intensity and to assess robustness to migration-related selection. See Appendix B.3 for further details.

We further merge time-varying local economic controls, including regional GDP per capita and consumer price indices, from the State Statistics Service of Ukraine.

3 Results: Ethnic and Linguistic Identity

3.1 Baseline Outcomes: Ethnicity and Language in the KIIS Survey

We begin by examining survey-based evidence on shifts in self-reported nationality and language use. The analysis relies on repeated cross-sectional data from the Kyiv International Institute of Sociology (KIIS), covering 2008–2024. We study how respondents' ethnic self-identification and interview language evolved after the outbreak of the two major phases of the Russia–Ukraine war: the 2014 conflict and the 2022 full-scale invasion. We examine whether these episodes consolidated a shared Ukrainian identity or generated polarization along ethnolinguistic lines.

3.1.1 Empirical Strategy

We proceed in two steps. First, we document raw time dynamics of reported nationality and interview language before and after each conflict episode. Second, we estimate difference-in-differences specifications that exploit pre-war cross-sectional variation in local ethnic composition. Specifically, we ask whether districts with larger pre-war Russian minorities exhibit weaker or stronger shifts toward Ukrainian identity following each conflict shock.

To estimate average post-war changes in identity outcomes, we first estimate the following pre–post specification:

$$\text{Identity}_{irt} = \beta \, \text{PostWar}_t + X'_{irt}\delta + \alpha_r + \varepsilon_{irt}. \tag{1}$$

We then examine heterogeneity in these shifts by local ethnic composition using a difference-in-differences specification:

$$\text{Identity}_{irt} = \beta \left(\text{PostWar}_t \times \text{EthnicRussians}_{r,2001}\right) + X'_{irt}\delta + \alpha_r + \gamma_t + \varepsilon_{irt}. \tag{2}$$

In both equations, Identity_{irt} denotes one of three outcomes: (i) an indicator for identifying as *only Ukrainian* (versus mixed or Russian) in the 2014 analysis; (ii) an

indicator for identifying as Ukrainian in the 2022 analysis; or (iii) an indicator for responding in Ukrainian (or mixed Ukrainian/Russian) rather than exclusively in Russian. In the 2014 analysis, PostWar_t equals one for survey waves after February 2014 and zero otherwise; in the 2022 analysis, it equals one for survey waves after February 2022 and zero otherwise. $\text{EthnicRussians}_{r,2001}$ is the pre-determined share of ethnic Russians in rayon r from the 2001 Census. X_{irt} denotes individual- and locality-level controls, while α_r and γ_t are rayon and survey-wave fixed effects, respectively. For the 2022 analysis, rayons are assigned based on respondents' pre-invasion place of residence as recorded in the survey. Standard errors are clustered at the rayon level.

In Equation (1), the coefficient β captures the average change in identity outcomes following each conflict episode. Identification in this specification relies on the assumption that, absent the conflict, average identity outcomes within rayons would have remained stable over time, conditional on controls and fixed effects.

In Equation (2), the coefficient β captures whether post-conflict identity shifts vary systematically with pre-war Russian ethnic presence. Identification relies on a parallel-trends assumption: absent the conflict, rayons with different pre-existing shares of ethnic Russians would have experienced similar changes in identity outcomes. We interpret positive estimates ($\beta > 0$) as evidence of identity consolidation and negative estimates ($\beta < 0$) as evidence of polarization.

Controls. We include several covariates to address compositional change and local economic conditions. To account for changes in respondent composition, we control for sex, age, and educational attainment, as well as an indicator for urban versus rural residence. In analyses of the post-2014 conflict, we additionally control for local displacement intensity, measured as the rayon-level ratio of internally displaced persons (IDPs) to population (see Appendix B.3 for details). For the 2022 event, we observe respondents' displacement status directly and include an indicator for individual displacement. To account for time-varying economic conditions, we merge region-by-time measures of real GDP per capita and the consumer price index (CPI)

to each respondent. Comparable economic controls are not available for the period surrounding the 2022 invasion and therefore cannot be included in those specifications.

3.1.2 The 2014 War

Ethnicity. We begin by estimating post-2014 changes in ethnic self-identification using Equation (1). Columns (1)–(2) of Panel A of Table 1 show a large aggregate shift toward Ukrainian ethnic identification. The unconditional post-2014 increase amounts to approximately 8.9 percentage points. Once individual- and rayon-level controls are included, the estimate declines to 3.7 percentage points but remains statistically significant, indicating that part of the aggregate shift is correlated with contemporaneous changes in local economic conditions.

We next turn to the difference-in-differences specification in Equation (2). Columns (1)–(2) of Panel B of Table 1 provide strong evidence of identity consolidation: the increase in exclusive Ukrainian identification is substantially larger in rayons with higher pre-war shares of ethnic Russians. The interaction term *Post 2014* $\times$ *Share of Ethnic Russians* is positive and highly significant. Quantitatively, moving a respondent from a district with 0% to a district with 25% ethnic Russians increases the probability of identifying as *only Ukrainian* after the 2014 conflict by about 14 percentage points.

Figure 4 presents event-study estimates from a wave-by-wave version of Equation (2). The estimates are flat before February 2014, rise thereafter, and remain elevated throughout the post-conflict period, with no evidence of differential pre-trends.

Finally, Table C.1 extends both the pre–post and difference-in-differences analyses to the full Likert scale of ethnic identity. The results show that the rise in exclusive Ukrainian identification comes at the expense of Russian and mixed identities.[6]

[6]These patterns are also visible in the raw data. Figure C.1 plots mean shares of reported ethnicity before and after the onset of the 2014 conflict, showing a ten percentage point increase in exclusive Ukrainian identification and corresponding declines in mixed and Russian identities. Figure 1 plots before–after changes in ethnic identification by pre-war Russian population share, with larger shifts in areas with higher Russian minority shares.

Language. Columns (3)–(4) of Panel A of Table 1 report analogous results for the language of interview. Following 2014, the share of respondents not replying in Russian and instead replying in Ukrainian or a Ukrainian–Russian mix increases by roughly two percentage points. However, once individual- and rayon-level controls are included, the aggregate post-2014 shift is no longer statistically significant.

In contrast, the difference-in-differences results for language choice mirror the ethnicity findings. Columns (3)–(4) of Panel B of Table 1 and Table C.2 show that the shift away from Russian is significantly stronger in rayons with higher pre-war shares of ethnic Russians. The interaction term *Post 2014* × *Share of Ethnic Russians* is positive and statistically significant, with stable magnitudes across specifications both with and without controls. Quantitatively, moving a respondent from a district with 0% to a district with 25% ethnic Russians increases the probability of not replying in Russian after the 2014 conflict by approximately 3.5 percentage points.[7]

Event-study estimates in Figure 6 display no differential pre-trends and a persistent increase after February 2014.

Taken together, the results so far reveal a divergence between identity and language adjustment in the immediate aftermath of the 2014 conflict: ethnic self-identification responds sharply, while language use adjusts more gradually. At the same time, both outcomes exhibit patterns consistent with national consolidation. The full-scale invasion of 2022 provides a natural setting to examine whether a larger and more persistent war shock may lead to broader linguistic realignment.

3.1.3 The 2022 War

Ethnicity. After the full-scale invasion, KIIS discontinued the five-category Likert question on ethnic identity. However, a direct question on nationality—allowing

[7] Figure C.2 reports raw before–after changes in interview language by pre-war Russian population share. The figure shows an increase in Ukrainian-language responses relative to mixed Surzhyk following the 2014 conflict, concentrated in rayons with lower Russian minority shares. In rayons with higher Russian minority shares, the adjustment occurs primarily through a shift from Russian to mixed Surzhyk rather than to exclusively Ukrainian.

respondents to identify as Ukrainian, Russian, or mixed Ukrainian–Russian—remains available, and we use it for the 2022 analysis.

Columns (1)–(2) of Panel A of Table 2 document a large aggregate shift toward Ukrainian national identification after February 2022, amounting to 5.6–5.9 percentage points. Columns (1)–(2) of Panel B of Table 2 show that this shift is significantly stronger in rayons with higher pre-war shares of ethnic Russians: the interaction term *Post 2022* $\times$ *Share of Ethnic Russians* is positive and highly significant, and remains robust to the inclusion of controls in column (2). Quantitatively, moving from a rayon with 0% ethnic Russians to one with 25% ethnic Russians in the 2001 Census is associated with an approximately 10 percentage-point higher probability of identifying as Ukrainian. Event-study estimates in Figure 5 are flat prior to early 2022 and rise sharply thereafter, with no evidence of reversion through 2024. Taken together, these patterns point to strong national consolidation rather than ethnic polarization following the 2022 invasion, consistent with the patterns observed after the 2014 conflict.[8]

Language. Figure 3 documents one of the most striking shifts in our data. Following February 2022, the share of respondents replying in Ukrainian or a Ukrainian-Russian mix increases by approximately 20 percentage points, mirrored by a comparable decline in Russian. In other words, roughly one-fifth of respondents switch the language of interview after the invasion. Columns (3)–(4) of Panel A of Table 2 confirm the magnitude of this shift in a regression framework, with and without controls.

Consistent with earlier findings, the shift is strongest in rayons with higher pre-war shares of ethnic Russians, reinforcing the interpretation of the invasion as inducing national consolidation rather than ethnic polarization. Columns (3)–(4) of Panel B of Table 2 confirm this heterogeneity pattern. Table C.3 further shows that the adjustment occurs primarily through movement from Russian to Ukrainian, rather than through mixed-language responses. The interaction term *Post 2022* $\times$ *Share of Ethnic Russians*

[8]Figure 2 reports raw mean changes in reported nationality before and after the invasion. The share identifying as Ukrainian increases by about seven percentage points, while mixed and Russian identities decline sharply; by 2022, only around 2.5% of respondents identify as Russian.

is large, positive, and robust to the inclusion of controls in column (4): moving from 0% to 25% ethnic Russians increases the probability of not replying in Russian by approximately 22 percentage points. Event-study estimates in Figure 7 display an immediate break at the onset of the war that remains elevated throughout the post-invasion period, with no evidence of differential pre-trends or subsequent reversal.

Taken together, the evidence from 2014 and 2022 points to a consistent pattern of national consolidation across conflict episodes, but with different margins of adjustment. Following the 2014 conflict, shifts are concentrated in ethnic self-identification, while changes in language use are more muted. In contrast, the 2022 invasion generates large and immediate adjustments along both dimensions, with particularly pronounced changes in language use. This contrast is consistent with ethnic identification adjusting more readily to political shocks, while language use—shaped by habits and coordination—responds more strongly once the shock is sufficiently large and persistent. We next address potential concerns related to survey responses and compositional change, and then examine whether similar patterns appear in measures of private linguistic behavior, including Twitter and Google search data.

3.1.4 Potential Concerns

Two concerns are salient for the above survey-based analysis: social desirability in survey responses and compositional change driven by migration and displacement.

Social desirability bias. If respondents became reluctant to report Russian ethnicity after 2014 or 2022, one would expect an increase in item nonresponse. Figure C.3 shows no meaningful post-2014 or post-2022 change in the share of "no reply" answers to the ethnicity question. This pattern is difficult to reconcile with an interpretation in which the main results are driven by increased refusal or concealment rather than genuine shifts in identity. In addition, Figure C.4 shows only a modest deterioration in attitudes toward Russian-speaking Ukrainians following the 2022 invasion—too small to account for the large changes in language use documented above. We further

corroborate the survey evidence using outcomes based on private behavior: post-2022 shifts also appear in Google search behavior and music consumption, which are less exposed to interviewer demand and are observed outside the survey setting.

Migration, refugees, and IDPs. Population movements could mechanically alter the composition of respondents within locations. We address this concern in several ways. First, our specifications control for local displacement intensity. For the 2014 conflict, we use the rayon-level ratio of internally displaced persons (IDPs) to population; for 2022, we control for individual displacement status and assign spatial fixed effects based on respondents' pre-invasion place of residence. Second, Figure C.5 examines heterogeneity by age and sex in 2022 and reveals similar effects across groups, including among military-age men, who were generally not permitted to leave Ukraine after the invasion. This substantially limits the scope for selective outmigration to explain the observed linguistic shifts. Third, we use Twitter data to estimate models with account fixed effects, so identification comes from within-account changes over time and cannot be driven by changes in the composition of users. The Twitter evidence closely mirrors the survey findings. Finally, surveys of Ukrainian refugees in the European Union indicate that approximately 70% report Ukrainian as their primary language of communication—substantially above the pre-invasion national average.[9]

3.2 Language of Tweets

We complement the survey-based evidence with high-frequency data from social media. Twitter offers a measure of public language use, enabling us to observe identity-related shifts within the same users. This approach thus complements the KIIS survey results by capturing how the choice between Ukrainian and Russian evolved in public online communication, rather than in a one-off interaction with a survey interviewer and making the analysis robust to population movements.

To estimate within-individual dynamics, we collapse tweets to the account-by-

[9] International Organization for Migration (2023), *Ukrainian Refugee Response in the European Union: Needs, Intentions, and Integration Challenges*. Migration Research Series No. 74.

month level and compute, for each account, the share of tweets written in Ukrainian in a given month. We then regress this monthly share on account fixed effects and month fixed effects. The inclusion of account fixed effects ensures that variation comes from within-account changes over time, so the estimated month effects trace shifts in the propensity to tweet in Ukrainian relative to January 2022. Importantly, this specification assigns equal weight to each account rather than to individual tweets, preventing highly active users from mechanically driving the results.[10]

Figure 8 plots the estimated month fixed. The series exhibits a gradual upward trend from 2012 through 2021, with no discrete break around the onset of hostilities in 2014—consistent with the absence of a major level shift in language use in the survey data. In contrast, the post-2022 shift in Ukrainian-language use is large and abrupt, on the order of 20 percentage points, and persists through the end of the sample. The timing and magnitude of this shift closely mirror the survey evidence.

Overall, the Twitter evidence corroborates the survey results and supports the interpretation of a genuine aggregate shift in linguistic practice after the 2022 invasion.[11] The move toward Ukrainian appears not only in survey interviews but also in public online communication, where language choice is observed repeatedly for the same users and is not mediated by an interviewer. This pattern indicates that the changes we document reflect broad and persistent adjustments in everyday language use rather than reporting context or sample composition.

3.3 Language of Google Searches

We next examine private linguistic behavior using Google search data. This analysis complements the survey and Twitter evidence by capturing revealed language choice in a setting that is largely private and therefore less susceptible to social desir-

[10] Formally, we estimate $\text{ShareUkr}_{at} = \alpha_a + \gamma_t + u_{at}$, where ShareUkr_{at} is the share of Ukrainian-language tweets posted by account a in month t, and α_a and γ_t denote account and month fixed effects, respectively. Standard errors are clustered at the account level.

[11] Due to changes in privacy policies and declining user preferences for sharing precise locations, the share of geotagged tweets from Ukraine fell from roughly 70% in 2014 to 7% in 2022 (Abramenko et al., 2024). This dramatic shift implies increasing selection in the sample of precisely geolocated tweets over time, rendering regional heterogeneity analysis unreliable.

ability, interviewer demand, or reputational concerns, providing a stringent test of conflict-induced identity change. Specifically, we track pairs of Ukrainian and Russian spellings of the same high-frequency words, allowing us to isolate language choice in search behavior rather than changes in what people search for (Appendix B.2.1).

To focus on changes associated with the 2022 invasion, we study Ukrainian versus Russian search activity after removing seasonality and extrapolating pre-invasion trends. Raw dynamics are shown in Figure C.6, which plots quarterly time effects from a regression of the Ukrainian-language search share on time dummies. These raw series exhibit both a pronounced seasonal component and a gradual pre-2022 upward trend, motivating a detrended analysis.

Our detrending procedure consists of three steps. First, we remove seasonality as well as word- and region-specific variation by residualizing the share of searches using Ukrainian spellings with respect to a full set of calendar-quarter by word by region fixed effects. Second, restricting attention to the pre-invasion period, we estimate word- and region-specific linear time trends in the resulting residuals and use the fitted values to construct predicted residual dynamics absent the invasion. Finally, we extrapolate these predicted trends into the post-invasion period and subtract them from the former residuals, yielding a series that captures deviations from both fixed effects and pre-invasion trends.[12] All regressions restrict attention to fully controlled regions, apply the weights described in Appendix B.2.1, and cluster standard errors at the word–region level.

Figure 9 presents these detrended time effects. Ukrainian-language searches increase sharply following the invasion: by July 2022, the share of searches in Ukrainian rises by 7.8 percentage points relative to the pre-invasion baseline and subsequently stabilizes between 8 and 10 percentage points over the rest of the sample period. Together with the survey and Twitter evidence, these results indicate a large and

[12] Formally, we first estimate $S_{tir} = \gamma_{qir} + \varepsilon_{tir}$, where S_{tir} denotes the share of searches using Ukrainian spelling for word i in region r at time t, occupied and partially occupied regions are excluded, and γ_{qir} are calendar-quarter by word by region fixed effects. We then retain $\hat{\varepsilon}_{tir}$. Then, using pre-2022 data, we estimate $\hat{\varepsilon}_{tir} = \beta_i \times t + \delta_r \times t + u_{tir}$, obtain and extrapolate $\widehat{\varepsilon}^{\text{pre}}_{tir}$, then define and plot $\tilde{\varepsilon}_{tir} = \hat{\varepsilon}_{tir} - \widehat{\varepsilon}^{\text{pre}}_{tir}$.

persistent shift toward Ukrainian language use that extends to private settings.[13]

4 Results: Cultural Shifts

We complement the survey- and social-media evidence on ethnicity and language with measures of cultural consumption. While survey responses capture stated identity and interview language reflects a salient communicative choice, consumption-related behaviors provide a different and arguably more demanding test of identity change: they are revealed-preference and often private outcomes that can involve switching costs or habit persistence. We therefore examine cultural markers that vary in their degree of visibility and commitment. Baby names represent a particularly costly and durable signal, reflecting long-run identity choices made at the household level. Cuisine and music, in contrast, capture day-to-day consumption and cultural attention, mostly private, allowing us to assess whether the post-war shift toward Ukrainian identity is accompanied by corresponding changes in cultural preferences. Together, these outcomes help establish whether war-induced identity change extends beyond self-reports and public communication to the domain of everyday consumption.

4.1 Baby names

We begin by studying whether baby naming patterns shift toward more Ukrainian names after the February 2022 full-scale invasion. Naming choices provide a complementary and durable measure of identity. Because names are public, highly persistent, and costly to reverse, shifts in naming patterns offer a useful signal of changes in ethnic and national identification (e.g., Fouka, 2019; Esposito et al., 2023).

Appendix Section B.2.3 describes the data and construction of the name Ukrainianness index, which assigns each name a time-invariant score based on its pre-war relative popularity in Ukraine versus Russia. As an illustration, names such as Bohdan

[13]The magnitude of the shift in Google search behavior is smaller than the corresponding change in survey-based language use, a pattern that is consistent with the presence of social desirability bias in self-reported language measures. We refrain from making quantitative claims, however, as this difference may also stem from Google search users not being a nationally representative sample of the Ukrainian population.

or Solomiya receive relatively high Ukrainianness scores, whereas names such as Yelysei or Yeseniia receive relatively lower scores.

We study how the *popularity* of a name co-moves with its Ukrainianness over time. The estimating equation is:

$$\text{Popular}_{rtn} = \sum_{s \neq 2021\text{ II}} \theta_s(\text{Ukrainianness}_n \times \mathbb{1}[t = s]) + \alpha_{rt} + \eta_{rn} + \varepsilon_{rtn}, \quad (3)$$

where Popular_{rtn} is an indicator equal to one if name n appears in the list of popular names in region r at time t.[14] The coefficients θ_s capture how the association between a name's Ukrainianness and its likelihood of appearing on the list in period s differs from the reference period, the second half of 2021. The specification includes region–time fixed effects α_{rt} and region–name fixed effects η_{rn}, so identification comes from changes over time in the relative popularity of more- versus less-Ukrainian names within the same region. Standard errors are two-way clustered by region and name.

Figure 10 plots the event-study coefficients θ_s. The estimates turn positive after 2022, indicating that names with higher Ukrainianness gained relative popularity following the full-scale invasion. We find no evidence of differential pre-trends. Estimating a simple pre-post version of Equation (3) yields the estimates reported in Table 3. The coefficient on Post-February 2022 $\times$ Ukrainianness is about 5.1 percentage points and is statistically significant at conventional levels. This implies that after the full-scale invasion, a fully Ukrainian name became about 5.1 p.p. more likely to appear on a region-period popularity list than a fully Russian name, corresponding to roughly 38% of the sample mean. The estimated effect is positive for both boys' and girls' names, with somewhat larger magnitudes for girls (6.1 p.p.) than for boys (4.0 p.p.). The coefficient for girls is statistically significant at conventional levels, while the estimate for boys is less precisely estimated.

Our main results are robust to alternative specifications that use average name

[14] On average, regional lists contain 28 popular names per period. Region–time fixed effects account for variation in the size of these lists across regions and periods.

Ukrainianness at the region–period level as the outcome, where we can explicitly control for population displacement and list composition (Appendix B.2.3; Figure B.4). Appendix Figure B.5 further shows that the post-2022 increase is broad-based across the name distribution, with larger gains for more Ukrainian names rather than being driven by a small set of outliers.

Taken together, these results indicate a broad-based shift in naming choices toward more Ukrainian names following the full-scale invasion, consistent with a strengthening of national identity that extends to durable and intergenerational decisions.

4.2 Cuisine

We next examine cuisine as an additional dimension of cultural identity change. Cuisine reflects everyday consumption choices based on revealed preferences rather than stated attitudes. We proxy interest in national Ukrainian cuisine using Google search activity for food and recipes, which captures private, routine consumption-related behavior at high frequency. While we are not aware of a comparable survey benchmark for Ukraine, survey evidence from other settings suggests that online recipe search is a mainstream activity: in a 2024 survey of U.S. consumers, 91% of respondents reported using online recipes (Chicory, 2024).

For each dish in our sample, we assign a time-invariant Ukrainianness score based on its relative pre-invasion popularity in Google Trends searches in Ukraine versus Russia. We then study how dish popularity evolves across regions over time as a function of this score. Appendix B.2.2 provides details on the data, index construction, and validation.

We estimate the following event-study specification, which allows the relationship between dish Ukrainianness and search popularity to vary by year:

$$\text{Popularity}_{irt} = \sum_{s \neq 2021} \theta_s \left(\text{Ukrainianness}_i \times \mathbb{1}[t = s]\right) + \alpha_{ir} + \delta_{rt} + \varepsilon_{irt}, \tag{4}$$

where Popularity_{irt} denotes Google Trends search intensity for dish i in oblast r and month t. In the baseline specification, we use the nonmodified Google Trends index

Popularity$^{raw}_{irt}$, normalized to its within-oblast maximum over 2012–2024. In an alternative specification, we use a standardized version, Popularity$^{std}_{irt}$, which is additionally demeaned and scaled to unit variance at the dish level to aid interpretation. The fixed effects α_{ir} absorb time-invariant differences in dish popularity across oblasts, while δ_{rt} capture common shocks to search activity by oblast-month. The coefficients β_y measure how the association between dish Ukrainianness and its popularity differs from its baseline 2021 level. Standard errors are two-way clustered by dish and oblast.

Figure 11 reports the yearly event-study estimates using the nonmodified index. Estimates are stable prior to 2022 and rise thereafter, indicating higher relative search intensity for more Ukrainian dishes within the same oblast-month. The post-2022 increase is amounting to roughly five points on the relative 0–100 Google Trends scale. Figure C.7 shows the same pattern using the standardized outcome, confirming that the result is not driven by dish-specific scaling. In standardized units, the post-2022 effect corresponds to approximately 0.2–0.3 standard deviations.

Finally, we assess whether the results are driven by a small number of high-share dishes. The cross-dish relationship between Ukrainianness and the war-period change in popularity is smooth, with no evidence that a handful of outliers account for the pattern. See Appendix B.2.2 for details.

Importantly, the estimated effect operates through searches for dishes rather than for branded products. The dish set consists of basic, non-branded food items, so the results are unlikely to be driven by the disappearance of Russian brands or by reduced availability of imported products from Russia.

Taken together, these results indicate that war-induced identity change is reflected not only in language use and self-identification, but also in everyday consumption choices. The reallocation toward more Ukrainian dishes suggests that identity shifts following the invasion altered revealed preferences in private settings, reinforcing the view that they can shape economic behavior beyond public or institutional contexts.

4.3 Popular Music and Spotify Charts

We next examine popular music, which provides a complementary consumption domain with richer dynamics. Music consumption provides a particularly informative lens through which to study cultural change. Unlike survey-based measures, listening behavior reflects private, revealed preferences and is therefore less susceptible to social-desirability bias. As an everyday consumption choice, music represents an understudied yet meaningful dimension of nation building. Examining shifts in music preferences allows us to assess whether changes in national identity extend beyond attitudes and into actual patterns of cultural consumption.

Spotify entered the Ukrainian market in July 2020 and has since become the leading music streaming platform in the country. By 2023, its market share reached 45% (Statista, 2025). Survey evidence confirms its broad adoption: 45% of Ukrainians report using Spotify, making it the most widely used paid streaming service (Suspilne Media, 2023). Usage is particularly concentrated among younger cohorts.

We use several subsamples derived from the datasets described in Appendix B.2.4: (i) the full chart sample (all Spotify Top-200 Ukraine song-week observations), (ii) chart observations restricted to songs released before 2022, (iii) chart observations for Russian artists matched to the war-stance classification, (iv) chart observations for the fixed list of Ukrainian artists, and (v) the Ukrainian-artist release panel used for supply-side analyses.

Chart-Level Language Dynamics We begin by documenting changes in the language composition of songs appearing in Spotify's weekly Top-200 charts in Ukraine. Figure 12 reports the share of chart entries by language, while Figure C.9 reports total weekly streams aggregated within each language group (Ukrainian, Russian, and Other).[15]

[15]The pronounced increase in the popularity of Russian-language music around the third quarter of 2021 appears to be associated with the release of the album *BANDANA I* by Russian artists on October 22, 2021. Tracks from the album gained substantial popularity and remained at the top of the charts for several weeks following the release.

Both measures show a sharp decline in the popularity of Russian-language music after the full-scale invasion. Before February 24, 2022, Russian-language songs accounted for roughly 60–70% of weekly chart entries, while Ukrainian-language songs rarely exceeded 5%. By May 2022, the share of Russian-language tracks had fallen to about 50%, while the share of Ukrainian-language tracks had risen above 20%. The shift continued thereafter: by early 2023, Ukrainian-language songs had overtaken Russian-language songs in chart presence, streaming volumes, and average chart performance. At the same time, we observe partial substitution toward songs in other languages (primarily English). This pattern is not driven by an increase in songs with missing lyrics,[16] whose chart share and streams remain relatively stable and generally below 10% throughout the period.

Demand-side reallocation holding new releases fixed. To assess whether the post-invasion shift in listening patterns is driven solely by the inflow of newly released Ukrainian-language music, we repeat the analysis restricting the chart sample to songs released before 2022. Figure C.10 shows that the main patterns remain: Russian-language songs decline sharply, while Ukrainian- and other-language songs gain relative popularity. The similarity between the full-sample and pre-2022 results indicates that the post-invasion shift is not solely a consequence of newly released Ukrainian-language music entering the charts. A mild seasonal pattern remains visible in the *Other* category, consistent with recurring popularity of holiday songs.

Mechanism (identity vs artist). A central question is whether the post-invasion shift in music consumption reflects a change in identity-linked language preferences, or instead a reallocation across artists (e.g., away from pro-war performers or toward Ukrainian artists regardless of song language). We address this in two complementary exercises. First, using the war-stance classification described in Appendix B.2.4, Figure C.12 shows that songs by pro-war artists and anti-war artists decline similarly from

[16] Lyrics may be missing because (i) a track is instrumental or (ii) lyrics have not yet been uploaded to Genius.

2022 onward. This pattern suggests that listeners increasingly avoid Russian-language music not only because of artists' war stance, but also independently of it.

Second, to separate language effects from artist effects, we fix the set of Ukrainian artists and compare chart outcomes by song language within that group. Figures C.13a and C.13b show a rapid reversal: within the same artist pool, the share of Russian-language songs in the charts falls from roughly 75% to below 20% within about a month after the invasion, while Ukrainian-language songs rise correspondingly. The same pattern remains when restricting to pre-war releases (Figures C.13c and C.13d), although the magnitude attenuates over time and Russian-language songs partly recover by 2024. Taken together, these results suggest that the observed reallocation in music consumption is not driven solely by artist nationality or political stance, but also reflects a broad shift in preferences over language itself.

Supply response. We then turn to the supply side and study changes in the language of new releases among Ukrainian artists. Figure C.11a reports the language composition of releases in the Ukrainian-artist catalogue panel. Before 2022, new releases were relatively stable in composition, with roughly 60% in Russian and about 20% in Ukrainian. After the invasion, this distribution shifts sharply: by the end of 2022Q2, Ukrainian-language releases surpass Russian-language releases, and by mid-2023 the pre-war pattern is close to reversed. Figure C.11c shows an even stronger shift among new artists (those whose first observed release appears after February 24, 2022): 80–90% of releases are in Ukrainian, while fewer than 5% are in Russian.

To test whether this pattern reflects within-artist language switching (rather than only entry/exit or compositional change across artists), we estimate the following artist fixed-effects regression separately for each language l:

$$share^{l}_{at} = \theta^{l}_{t} + \alpha_{a} + \varepsilon^{l}_{at}. \tag{5}$$

Here, $share^{l}_{at}$ denotes the share of releases by artist a in quarter t that are in language

l. The specification includes language-specific quarter fixed effects θ_t^l and artist fixed effects α_a. Standard errors are clustered at the artist level.

Figure C.11d reports the resulting quarter effects separately by language. Before February 24, 2022, the estimated effects are small and statistically indistinguishable from zero. After the invasion, a clear divergence emerges: within one year, the probability of releasing a song in Ukrainian increases by about 20 percentage points, while the probability of releasing a song in Russian declines by more than 20 percentage points. These estimates indicate that the supply response reflects substantial within-artist reorientation in language choice.

4.3.1 Concerns

Recommendations and platform promotion. A natural concern is whether changes in Spotify's recommendation system could have contributed to the observed rise in Ukrainian-language music. While anecdotal user reports suggest that Ukrainian songs may have appeared more frequently in recommendations after 2022, we are not aware of systematic evidence that Spotify altered its recommendation algorithms specifically to favor Ukrainian-language content. Spotify did, however, undertake visible editorial and promotional initiatives in support of Ukraine (including fundraising campaigns, curated playlists, and artist features) (Spotify Newsroom, 2022, 2023; Suspilne Media, 2024). These interventions may have increased the exposure of Ukrainian music, both directly through editorial placement and indirectly through recommendation systems that respond to aggregate listening behavior. We view such amplification as part of the broader post-invasion consumption environment rather than evidence that the documented patterns are mechanically generated by algorithmic reweighting alone.

Coverage in occupied and annexed regions. A second concern is whether conflict-related disruptions in Eastern and occupied regions changed the composition of Spotify users in ways that could mechanically generate the observed language shift. Although Spotify access may have been intermittent in partially occupied areas and

internet routing through Russian infrastructure may have affected availability (Spotify Community, 2022, 2023; Greenberg, 2022; BBC News, 2022), several pieces of evidence suggest that these factors are unlikely to fully account for our results. First, even prior to the full-scale invasion, Donetsk, Luhansk, and Crimea exhibited the lowest Spotify-related search interest within Ukraine (Table C.4). Second, territories occupied after 2022 represent no more than 9.6% of the population living in areas controlled by Ukraine as of January 1, 2022: summing oblast-level population estimates for the subsequently occupied territories yields 3.59 million out of 37.44 million (State Statistics Service of Ukraine, 2022). This calculation does not incorporate the large outflow from war zones—estimated at 2.83 million displaced persons (International Organization for Migration (IOM), 2025)—which would further reduce the weight of these areas in national listening aggregates. Third, the main patterns are not driven solely by a decline in Russian-language streams; Ukrainian-language tracks also rise sharply in absolute streaming volumes and account for a growing share of total streams (Figure C.9). Finally, while population movements could matter if emigrants were disproportionately Russian-speaking, available evidence does not support a shift large enough to overturn the magnitude and persistence of the observed reallocation.

5 Conclusion

This paper studies how large-scale external conflict reshapes national identity in a historically heterogeneous society. Using Ukraine as a setting, we examine how two major conflict shocks—the onset of hostilities in 2014 and Russia's full-scale invasion in 2022—affected ethnic identification, language use, and cultural behavior. We combine nationally representative surveys with digital traces of public and private behavior, allowing us to track identity change across multiple domains and levels of publicness.

We find consistent evidence of national consolidation rather than polarization. Both conflict episodes generate sharp and persistent shifts toward Ukrainian ethnic and linguistic identity, with the largest changes occurring precisely in regions that were historically more ethnolinguistically Russian. These patterns are difficult to

reconcile with narratives emphasizing ethnic fragmentation or backlash. Instead, they suggest that external threat can strengthen national identification even in societies with substantial pre-existing cultural and linguistic ties to the aggressor.

The adjustments we document extend well beyond stated identity. Following the 2022 invasion, language use shifts sharply toward Ukrainian not only in survey interviews, but also in public communication on social media and in private Google searches. Cultural consumption responds in parallel: Ukrainians increasingly favor Ukrainian-language music, Ukrainian cuisine, and traditionally Ukrainian baby names. These behaviors span margins with very different adjustment costs. Baby names represent durable, high-commitment stock decisions that are costly to reverse, while music and cuisine reflect everyday, low-stakes consumption flows. Yet even along these lower-commitment margins, the shifts are large and persistent. Taken together, this pattern strengthens the interpretation that the observed changes reflect genuine identity transformation rather than short-lived symbolic responses.

Methodologically, the paper contributes to the literature on conflict and nation-building by directly testing consolidation versus polarization within the same empirical setting. By combining survey data with high-frequency digital traces and administrative records, we also demonstrate how integrating multiple data sources can overcome long-standing concerns about social desirability, compositional change, and behavioral relevance in the study of identity.

More broadly, the results highlight the malleability of national identity in response to external violence. While identity is often viewed as slow-moving and deeply rooted, our evidence shows that large-scale conflict can generate rapid, durable, and behaviorally meaningful change—even in settings where fragmentation might have been expected ex ante. These findings have implications for theories of nation-building, for understanding collective resilience under threat, and for the long-run political and economic consequences of war.

Future research could examine the persistence of these identity shifts beyond the conflict horizon, their intergenerational transmission, and their role in shaping collec-

tive action, compliance, and societal resilience during and after large-scale conflict. More generally, the Ukrainian case underscores the importance of studying identity not only as a belief or attitude, but as a set of everyday choices that shape—and are shaped by—major political shocks.

References

Abramenko, S., V. Korovkin, and A. Makarin (2024). Social Media as an Identity Barometer: Evidence From the Russia-Ukraine War. AEA Papers and Proceedings 114, 70–74.

Aghion, P., X. Jaravel, T. Persson, and D. Rouzet (2019). Education and Military Rivalry. Journal of the European Economic Association 17(2), 376–412.

Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. Quarterly Journal of Economics 115(3), 715–753.

Akerlof, R. (2016). "We Thinking" and Its Consequences. American Economic Review: Papers and Proceedings 106(5), 415–419.

Alesina, A., P. Giuliano, and B. Reich (2021). Nation-Building and Education. The Economic Journal 131(638), 2273–2303.

Alesina, A., B. Reich, and A. Riboni (2020). Nation-Building, Nationalism, and Wars. Journal of Economic Growth 25(4), 381–430.

Almagro, M. and D. Andrés-Cerezo (2020). The Construction of National Identities. Theoretical Economics 15, 763–810.

Ananyev, M. and M. Poyker (2023). Identity and Conflict: Evidence from Tuareg Rebellion in Mali. World Development 161, 106108.

Anderson, B. (1983). Imagined Communities: Reflections on the Origin and Spread of Nationalism. London: Verso.

Arel, D. and V. Khmelko (1996). The Russian factor and territorial polarization in Ukraine. The Harriman Review 9(1-2), 81–91.

Assouad, L. (2025). Leadership and Nation-Building. Working Paper, London School of Economics.

Atkin, D., E. Colson-Sihra, and M. Shayo (2021). How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption. Journal of Political Economy 129(12), 3460–3508.

Barrington, L. W. (2018). Understanding Identity in Ukraine – and Elsewhere. Post-Soviet Affairs 34(2-3), 179–182.

Bauer, M., C. Blattman, J. Chytilová, J. Henrich, E. Miguel, and T. Mitts (2016). Can war foster cooperation? Journal of Economic Perspectives 30(3), 249–274.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2019). Unity in Diversity? How Intergroup Contact Can Foster Nation Building. American Economic Review 109(11), 3978–4025.

Bazzi, S., M. Hilmy, and B. Marx (2025). Religion, Education, and the State. Review of Economic Studies, rdaf071.

BBC News (2022). Spotify closes russian office and restricts content. https://www.bbc.com/news/business-60881567. Accessed: 27 August 2025.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. Quarterly Journal of Economics 126(2), 805–855.

Boulègue, M., J. Bronk, N. Bugayova, M. Clark, R. Häggblom, M. Konaev, and P. Luzin (2023). Understanding Russia's Military Vulnerabilities. Chatham House. Policy brief, Royal Institute of International Affairs.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? Journal of Social Issues 55(3), 429–444.

Büyükeren, B. (2025). Remaking the State from Below: Local Government Formation in Ukraine. Available at SSRN 5874582.

Bækken, H. (2024). Forging frontline russians: Militarized patriotism and identity formation in occupied ukraine. Communist and Post-Communist Studies 58(1).

Cannella, M., A. Makarin, and R. Pique (2024). The Political Legacy of Nazi Annexation. The Economic Journal 134(661), 1721–1759.

Carlitz, R., A. Morjaria, J. Mueller, and P. Osafo-Kwaako (2025). State Building in a Diverse Society. Review of Economic Studies 92(6), 3704–3740.

CFR (2024). Ukraine conflict tracker. https://www.cfr.org/global-conflict-tracker/conflict/conflict-ukraine.

Chicory (2024). The State of Online Recipes. Chicory. Accessed: 2026-02-25.

Clots-Figueras, I. and P. Masella (2013). Education, language and identity. Economic Journal 123(570), F332–F364.

D'Anieri, P. (2023). Ukraine and Russia: From Civilized Divorce to Uncivil War (2nd ed.). Cambridge: Cambridge University Press.

Darden, K. and H. Mylonas (2016). Threats to territorial integrity, national mass schooling, and linguistic commonality. Comparative Political Studies 49(11), 1446–1479.

Dehdari, S. H. and K. Gehring (2022). The origins of common identity: Evidence from alsace-lorraine. American Economic Journal: Applied Economics 14(1), 261–292.

Dell, M. and P. Querubin (2018). Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies. The Quarterly Journal of Economics 133(2), 701–764.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya (2014). Cross-border media and nationalism: Evidence from serbian radio in croatia. American Economic Journal: Applied Economics 6(3), 103–132.

Depetris-Chauvin, E., R. Durante, and F. Campante (2020). Building Nations through Shared Experiences: Evidence from African Football. American Economic Review 110(5), 1572–1602.

Esposito, E., T. Rotesi, A. Saia, and M. Thoenig (2023). Reconciliation Narratives. The American Economic Review 113(6), 1461–1504.

Federle, J., A. Meier, G. J. Müller, W. Mutschler, and M. Schularick (2024). The Price of War. Technical report, Kiel Working Paper.

Fouka, V. (2019). How Do Immigrants Respond to Discrimination? The Case of Germans in the US During World War I. American Political Science Review 113(2), 405–422.

Fouka, V. (2020). Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. Review of Economic Studies 87(1), 204–239.

Fritsche, I., M. Moya, M. Bukowski, P. Jugert, S. de Lemus, O. Decker, I. Valor-Segura, and G. Navarro-Carrillo (2017). The great recession and group-based control: Converting personal helplessness into social class in-group trust and collective action. Journal of Social Issues 73(1), 117–137.

Gehring, K. (2021). Can External Threats Foster a European Union Identity? Evidence from Russia's Invasion of Ukraine. Economic Journal 132(644), 1489–1516.

Gellner, E. (1983). Nations and Nationalism. Oxford: Blackwell.

Genius (2025). Genius api. https://docs.genius.com/. Accessed: 2025-08-22.

Goldhahn, D., T. Eckart, and U. Quasthoff (2012). Building large monolingual dictionaries at the leipzig corpora collection: From 100 to 200 languages. In Proceedings of the 8th International Language Resources and Evaluation (LREC'12), pp. 759–765. European Language Resources Association (ELRA).

Greenberg, A. (2022). Russia's quiet takeover of ukraine's internet. https://www.wired.com/story/ukraine-russia-internet-takeover/. Accessed: 27 August 2025.

Harm Adema, J. A., C. G. Aksoy, Y. Giesing, and P. Poutvaara (2024). The Effect of Conflict on Ukrainian Refugees' Return and Integration. SSRN Electronic Journal.

Helms, B., S. S. Pandya, and R. Venkatesan (2025). War on aisle 5: Casualties, national identity, and consumer behavior. Journal of Conflict Resolution 69(2-3), 434–462.

Hughes, J. (2011). State violence in the origins of nationalism: British counterinsurgency and the rebirth of irish nationalism, 1969–1972. In J. A. Hall and S. Malešević (Eds.), Nationalism and War, pp. 96–122. Cambridge University Press.

Hutchinson, J. (2017). Nationalism and War. Oxford: Oxford University Press.

International Organization for Migration (IOM) (2025, January). Ukraine internal displacement report, general population survey, round 19. Technical report, IOM. Accessed: 8 September 2025.

Jha, S. (2013). Trade, Institutions, and Ethnic Yolerance: Evidence from South Asia. American Political Science Review 107(4), 806–832.

Jia, R. and T. Persson (2021). Choosing ethnicity: The interplay between individual and social motives. Journal of the European Economic Association 19(5), 2487–2523.

Kochnev, A. (2019). Dying Light: War and Trade of the Separatist-Controlled Areas of Ukraine. Available at SSRN 3579099.

Korovkin, V. and A. Makarin (2023). Conflict and Intergroup Trade: Evidence from the 2014 Russia-Ukraine Crisis. American Economic Review 113(1), 34–70.

Korovkin, V., A. Makarin, and Y. Miyauchi (2025). Supply Chain Disruption and Reorganization: Theory and Evidence from Ukraine's War. Review of Economic Studies, forthcoming.

Kranton, R. E. (2016). Identity economics 2016: Where do social distinctions and norms come from? American Economic Review: Papers and Proceedings 106(5), 405–409.

Kravtsiv, B. and V. Kubijovyc (1993). Russification. In Encyclopedia of Ukraine. University of Toronto Press.

Kulyk, V. (2016). National Identity in Ukraine: Impact of Euromaidan and the War. Europe-Asia Studies 68(4), 588–608.

Kulyk, V. (2018). Shedding russianness, recasting ukrainianness: The post-euromaidan dynamics of ethnonational identifications in ukraine. Post-Soviet Affairs 34(2-3), 119–138.

Kulyk, V. (2024). Language shift in time of war: The abandonment of russian in ukraine. Post-Soviet Affairs 40(3), 159–174.

Kulyk, V. and H. E. Hale (2022). Imperfect measures of dynamic identities: The changing impact of ethnolinguistic characteristics on political attitudes in ukraine. Nations and Nationalism 28(3), 841–860.

Kworb (2025). Spotify chart history – ukraine weekly totals. https://kworb.net/spotify/country/ua_weekly_totals.html. Accessed: 2025-08-22.

Leipzig Corpora Collection (2025). Leipzig corpora collection. Accessed: 1 September 2025.

Magocsi, P. R. (1996). A History of Ukraine (2nd ed.). Toronto: University of Toronto Press.

Mann, M. (1993). The Sources of Social Power, Volume 2: The Rise of Classes and Nation–States, 1760–1914. Cambridge: Cambridge University Press.

Markevich, A., N. Naumenko, and N. Qian (2025). The Causes of Ukrainian Famine Mortality, 1932–33. Review of Economic Studies 92(5), 3276–3305.

Mayer, T., I. Méjean, and M. Thoenig (2026). Can Sanctions Deter Wars? The Russia-Ukraine Case.

Metzger, M. M., R. Bonneau, J. Nagler, and J. A. Tucker (2016). Tweeting Identity? Ukrainian, Russian, and #Euromaidan. Journal of Comparative Economics 44(1), 16–40.

Monde, L. (2024). Ukrainian as a language of resistance: "when i hear russian, it makes me want to vomit". https://www.lemonde.fr/en/international/article/2024/02/26/ukrainian-as-a-language-of-resistance-when-i-hear-russian-it-makes-me-want-to-vomit_6558620_4.html. Accessed: insert date.

Mueller, J. E. (1970). Presidential Popularity from Truman to Johnson. American Political Science Review 64(1), 18–34.

Nardotto, M. and S. Sequeira (2025). Identity, media and consumer behavior. Review of Economic Studies. Forthcoming.

NUAM Club (2025). Nuam club website. https://www.nuam.club/. Accessed: 2025-08-22.

Osypenko, O. (2025, September). Wartime resurgence: Surzhyk, a blend of ukrainian and russian once stigmatised as a marker of backwardness, is enjoying an unexpected comeback. Accessed: insert date.

Pandya, S. S. and R. Venkatesan (2016). French roast: Consumer responses to international conflict—evidence from supermarket scanner data. Review of Economics and Statistics 98(1), 73–88.

Plokhy, S. (2015). The Gates of Europe: A History of Ukraine. New York: Basic Books.

Pop-Eleches, G. and G. B. Robertson (2018). Identity and political preferences in ukraine – before and after the euromaidan. Post-Soviet Affairs 34(2-3), 107–118.

Rapawy, S. (1997). Ethnic reidentification in ukraine. IPC Staff Paper 90, U.S. Bureau of the Census, International Programs Center, Washington, D.C.

Rohner, D. and E. Zhuravskaya (2023). Nation building: Big lessons from successes and failures. CEPR Press.

Rohner, D. and E. Zhuravskaya (2025). The Economics of Nation-Building: Methodological Tool Kit and Policy Lessons. Annual Review of Economics 17.

Rui J. P. de Figueiredo, J. and Z. Elkins (2003). Are patriots bigots? an inquiry into the vices of in-group pride. American Journal of Political Science 47(1), 171–188.

Sambanis, N. and A. H.-Y. Lee (2023). Does external threat bring the nation together? Working Paper, Penn Identity and Conflict (PIC) Lab, University of Pennsylvania.

Sambanis, N. and M. Shayo (2013). Social identification and ethnic conflict. American Political Science Review 107(2), 294–325.

Sambanis, N., S. Skaperdas, and W. C. Wohlforth (2015, May). Nation-Building through War. American Political Science Review 109(2), 279–296.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. American Political Science Review 103(2), 147–174.

Shayo, M. and A. Zussman (2011). Judicial Ingroup Bias in the Shadow of Terrorism. The Quarterly Journal of Economics 126(3), 1447–1484.

Smith, A. D. (1981). War and ethnicity: The role of warfare in the formation, self-images and cohesion of ethnic communities. Ethnic and Racial Studies 4(4), 375–397.

Smith, A. D. (1986). State-making and nation-building. In J. A. Hall (Ed.), States in History, pp. 228–263. Oxford: Basil Blackwell.

Smith, A. D. (1998). Nationalism and Modernism: A Critical Survey of Recent Theories of Nations and Nationalism. London and New York: Routledge.

Spotify (2025a). Spotify charts. https://charts.spotify.com/home. Accessed: 2025-08-22.

Spotify (2025b). Spotify playlist: Ukrainian music. https://open.spotify.com/playlist/4vjrpNQwniulRL1EMKX9Pq?si=zXNHlWIkSf2ulaq68apYjQ&pi=MjDxQJ9FRfOIO. Accessed: 2025-08-22.

Spotify (2025c). Spotify web api documentation. https://developer.spotify.com/documentation/web-api. Accessed: 2025-08-22.

Spotify Community (2022). Spotify app & website not working in ukraine. https://community.spotify.com/t5/Other-Podcasts-Partners-etc/Spotify-App-amp-Website-not-working-in-Ukraine/td-p/5358653. Accessed: 27 August 2025.

Spotify Community (2023). Spotify is unavailable in donetsk region, ukraine without using vpn. https://community.spotify.com/t5/Accounts/Spotify-is-unavailable-in-Donetsk-region-Ukraine-without-using/td-p/6029089. Accessed: 27 August 2025.

Spotify Newsroom (2022). Spotify mobilizes fans, artists, and our "music for ukraine" playlist ahead of ukrainian independence day. https://newsroom.spotify.com/2022-08-18/spotify-mobilizes-fans-artists-and-our-music-for-ukraine-playlist-ahead-of-ukrainian Accessed: 27 August 2025.

Spotify Newsroom (2023). Jamala kicks off spotify equal's ukrainian launch. https://newsroom.spotify.com/2023-03-10/jamala-kicks-off-spotify-equals-ukrainian-launch-by-sharing-her-unique-blend-of-soul Accessed: 27 August 2025.

Stars About War Project (2025). Stars about war. https://starsaboutwar.in.ua/. Accessed: 2025-08-22.

State Statistics Service of Ukraine (2022). Population estimates as of january 1, 2022. Technical report, State Statistics Service of Ukraine. Accessed: 8 September 2025.

Statista (2025). Digital music — ukraine: Market data and outlook. https://www.statista.com/outlook/amo/media/music-radio-podcasts/digital-music/ukraine. Accessed: 27 August 2025.

Subtelny, O. (2009). Ukraine: A History (4th ed.). Toronto: University of Toronto Press.

Suspilne Media (2023). What and how ukrainians listen: results of a study of musical preferences in 2023.

Suspilne Media (2024). Kalush orchestra released its second project for spotify singles. https://suspilne.media/culture/529097-kalush-orchestra-vipustiv-uze-drugij-proekt-dla-spotify-singles/. Accessed: 27 August 2025.

Tilly, C. (1975). The Formation of National States in Western Europe. Princeton, NJ: Princeton University Press.

UN High Commissioner for Refugees (2025). Ukraine emergency response. https://www.unrefugees.org/emergencies/ukraine/.

Weisel, O. and R. Zultan (2016). Social motives in intergroup conflict: Group identity and perceived target of threat. European Economic Review 90, 122–133.

Wikipedia contributors (2025). Category: Singers of Ukraine.

Wimmer, A. (2012). Waves of War: Nationalism, State Formation, and Ethnic Exclusion in the Modern World. Cambridge: Cambridge University Press.

Tables

Table 1: The Effect of 2014 War on the National Identity of Ukrainian Citizens

	Identified As Ukrainian Only (%)		Replied Not In Russian (%)	
	(1)	(2)	(3)	(4)
Panel A: Post-war shift				
Post-February 2014	8.896***	3.753**	2.122**	0.756
	(1.309)	(1.527)	(0.923)	(2.042)
Dependent variable mean	79.757	79.782	60.780	60.761
Dependent variable st. dev.	40.181	40.163	48.824	48.829
Observations	87,751	87,272	95,666	95,117
Controls		✓		✓
Rayon FE	✓	✓	✓	✓
Wave FE				
Panel B: Difference-in-differences				
Post-February 2014 × share of ethnic Russians	71.038***	55.669***	16.215***	13.914**
	(9.090)	(11.594)	(4.525)	(5.808)
Dependent variable mean	79.757	79.782	60.780	60.761
Dependent variable st. dev.	40.181	40.163	48.824	48.829
Observations	87,751	87,272	95,666	95,117
Controls		✓		✓
Rayon FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓

Notes: This table reports regression estimates of ethnic and linguistic identity outcomes on a post-February 2014 indicator and its interaction with pre-war local ethnic composition, testing the polarization hypothesis. Dependent variables are indicator variables multiplied by 100; coefficients are in percentage points. The dependent variables are *Identified as Ukrainian only* (columns 1–2), an indicator equal to one if the respondent identifies exclusively as Ukrainian (relative to Russian or mixed identities), and *Replied not in Russian* (columns 3–4), an indicator equal to one if the respondent's interview language was not Russian (Ukrainian or mixed Ukrainian–Russian). Panel A reports coefficients on *Post-February 2014*, an indicator equal to one for interviews conducted after February 2014. Panel B reports coefficients on *Post-February 2014 × share of ethnic Russians*, where the share is measured at the district level (pre-2020 rayons and cities of regional significance) using the 2001 Census. All specifications include district fixed effects; Panel B specifications additionally include survey-wave fixed effects. Panel A specifications omit wave fixed effects because they would absorb the post-February 2014 indicator. Columns (2) and (4) additionally control for sex, age, education, an urban/rural indicator, district-level internally displaced persons inflows per population, regional GDP per capita, and the CPI. Standard errors in parentheses are clustered at the district level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: The Effect of 2022 Invasion on the National Identity of Ukrainian Citizens

	Identified As Ukrainian Only (%)		Replied Not In Russian (%)	
	(1)	(2)	(3)	(4)
Panel A: Post-war shift				
Post-February 2022	5.867***	5.640***	21.180***	19.708***
	(0.682)	(0.675)	(1.809)	(1.837)
Dependent variable mean	92.254	92.251	69.675	69.652
Dependent variable st. dev.	26.732	26.736	45.967	45.977
Observations	50,673	50,512	48,976	48,744
Controls		✓		✓
Rayon FE	✓	✓	✓	✓
Panel B: Difference-in-differences				
Post-February 2022 × share of ethnic Russians	41.641***	39.226***	95.036***	88.406***
	(3.391)	(3.006)	(8.354)	(9.321)
Dependent variable mean	92.254	92.251	69.675	69.652
Dependent variable st. dev.	26.732	26.736	45.967	45.977
Observations	50,673	50,512	48,976	48,744
Controls		✓		✓
Rayon FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓

Notes: This table reports regression estimates of ethnic and linguistic identity outcomes on a post-February 2022 indicator and its interaction with pre-war local ethnic composition, testing the polarization hypothesis. Dependent variables are indicator variables multiplied by 100; coefficients are in percentage points. The dependent variables are *Identified as Ukrainian only* (columns 1–2), an indicator equal to one if the respondent identifies exclusively as Ukrainian (relative to Russian or mixed identities), and *Replied not in Russian* (columns 3–4), an indicator equal to one if the respondent's interview language was not Russian (Ukrainian or mixed Ukrainian–Russian). Panel A reports coefficients on *Post-February 2022*, an indicator equal to one for interviews conducted after February 2022. Panel B reports coefficients on *Post-February 2022 × share of ethnic Russians*, where the share is measured at the district level (pre-2020 rayons and cities of regional significance) using the 2001 Census. All specifications include district fixed effects (post-2020 rayons); Panel B specifications additionally include survey-wave fixed effects. Panel A specifications omit wave fixed effects because they would absorb the post-February 2022 indicator. For internally displaced respondents, the district is defined as the pre-February 2022 district of residence. Columns (2) and (4) additionally control for sex, age, education, an urban/rural indicator, pre-2020 district level share of ethnic Russians, and an indicator for being internally displaced interacted with the survey-wave identifier. Standard errors in parentheses are clustered at the post-2020 district level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Baby-Name Popularity and Ukrainianness

	Popular name indicator			
	(1)	(2)	(3)	(4)
Sample	All	All	Boys	Girls
Post-February 2022 × Ukrainianness	5.016**	5.148**	4.050	6.130***
	(1.849)	(1.874)	(2.654)	(2.124)
Dependent variable mean	13.593	13.593	14.774	12.528
Dependent variable st. dev.	34.271	34.271	35.485	33.104
Observations	83,134	83,134	39,400	43,734
Region × Time FE		✓	✓	✓
Region × Name FE		✓	✓	✓

Notes: This table reports regression estimates of the post–February 2022 change in the association between a baby name's popularity and its pre-war *Ukrainianness*. The unit of observation is name $i\times$ region $r\times$ half-year t. The dependent variable, $Popular_{irt}$, is an indicator equal to one if name i appears in the Ministry of Justice list of most popular baby names in region r and period (half-year/year) t; it is multiplied by 100, so coefficients are in percentage points. The key regressor is the interaction $Post_t \times I_i^U$, where $I_i^U \equiv \frac{U_i}{U_i+R_i}$ is a name-level index computed from pre-war name frequencies: U_i is the pre-war share of name i among births in Ukraine and R_i is the analogous share in Russia, taken from forebears.io. Columns report the coefficient on $Post_t \times I_i^U$ for the full sample (columns 1–2), boys (column 3), and girls (column 4). Column (1) includes no fixed effects. Columns (2)–(4) add region×period fixed effects and region×name fixed effects. Standard errors in parentheses are two-way clustered by region and by name. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figures

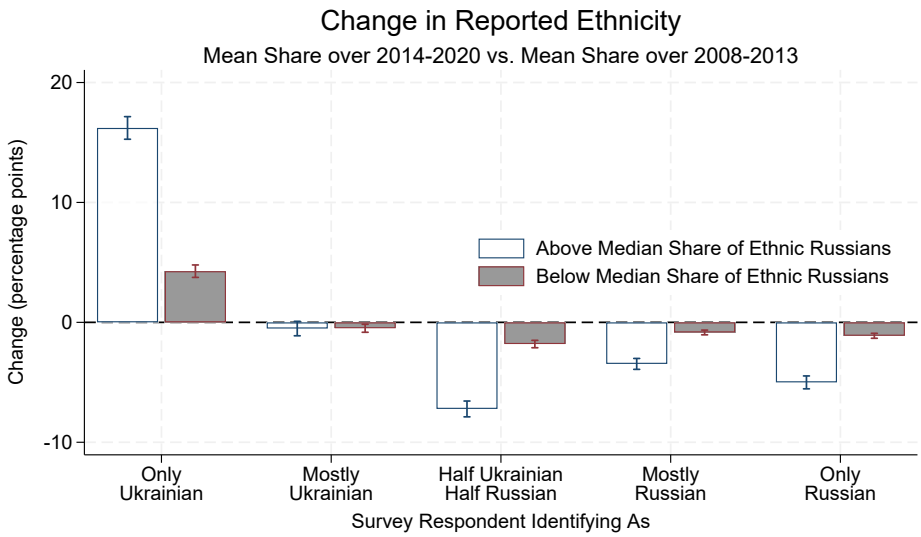


Figure 1: Change in Ethnic Identity Post 2014 War, by Pre-War Ethnic Composition

Notes: This figure plots changes in the mean share of respondents in each self-reported ethnicity category in the KIIS survey between the post-war period (March 2014–2020) and the pre-war period (2008–February 2014). Categories are: only Ukrainian, mostly Ukrainian, half Ukrainian–half Russian, mostly Russian, and only Russian. The sample excludes observations from Crimea and from districts that were fully or partially occupied after 2014. Bars report $\Delta\bar{x} = \bar{x}_{2014\text{–}2020} - \bar{x}_{2008\text{–}2013}$. White bars correspond to respondents in districts with an above-median pre-war share of ethnic Russians measured using the 2001 Census; gray bars correspond to respondents in districts with a below-median share. Error bars show 95% confidence intervals for the difference in period means.

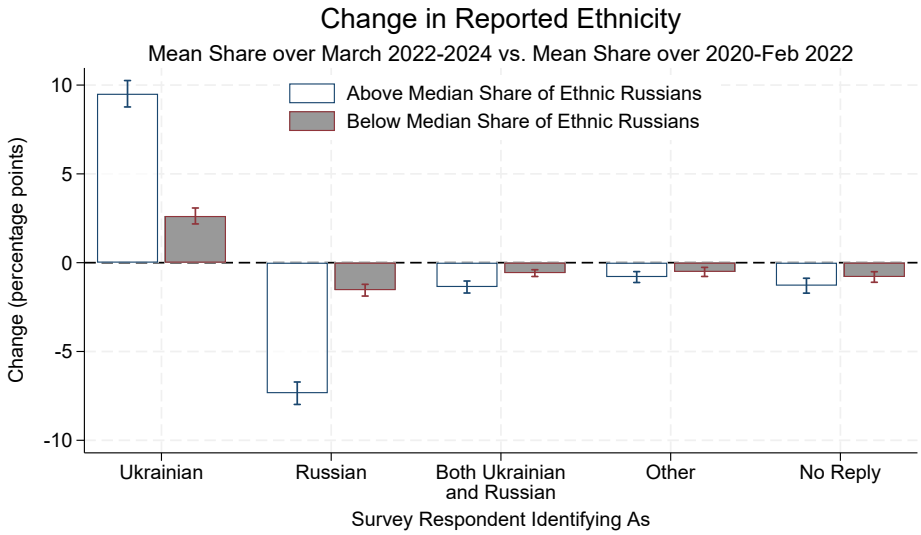


Figure 2: Change in Ethnic Identity Post 2022 Invasion, by Pre-War Ethnic Composition

Notes: This figure plots the change in the mean share of respondents in each self-reported ethnicity category in the KIIS survey between the post-invasion period (March 2022–2024) and the pre-invasion period (2020–February 2022). Categories are: Ukrainian, Russian, both Ukrainian and Russian, other ethnicity, and no reply. Bars report $\Delta\bar{x} = \bar{x}_{\text{Mar 2022–2024}} - \bar{x}_{\text{2020–Feb 2022}}$. White bars correspond to respondents in districts with an above-median pre-war share of ethnic Russians measured using the 2001 Census; gray bars correspond to respondents in districts with a below-median share. Error bars show 95% confidence intervals computed from the difference in period means.

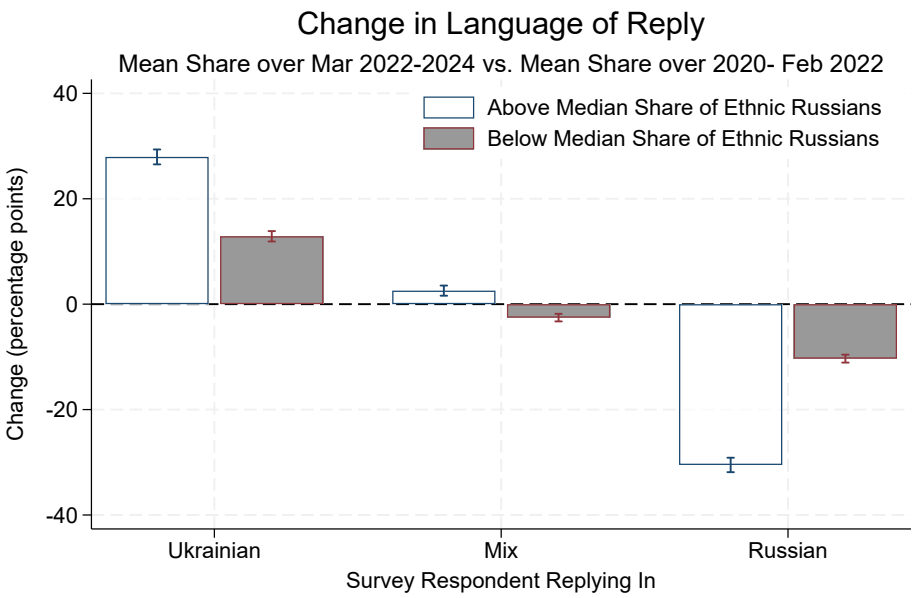


Figure 3: Change in Language of Reply Post 2022 Invasion, by Pre-War Ethnic Shares

Notes: This figure plots the change in the mean share of KIIS respondents replying in Ukrainian, mixed Ukrainian–Russian, and Russian between the post-invasion period (March 2022–2024) and the pre-invasion period (2020–February 2022). Bars report $\Delta\bar{x} = \bar{x}_{\text{Mar 2022–2024}} - \bar{x}_{\text{2020–Feb 2022}}$, expressed in percentage points. White bars correspond to respondents in districts with an above-median pre-war share of ethnic Russians measured using the 2001 Census; gray bars correspond to respondents in districts with a below-median share. Error bars show 95% confidence intervals computed from the difference in period means.

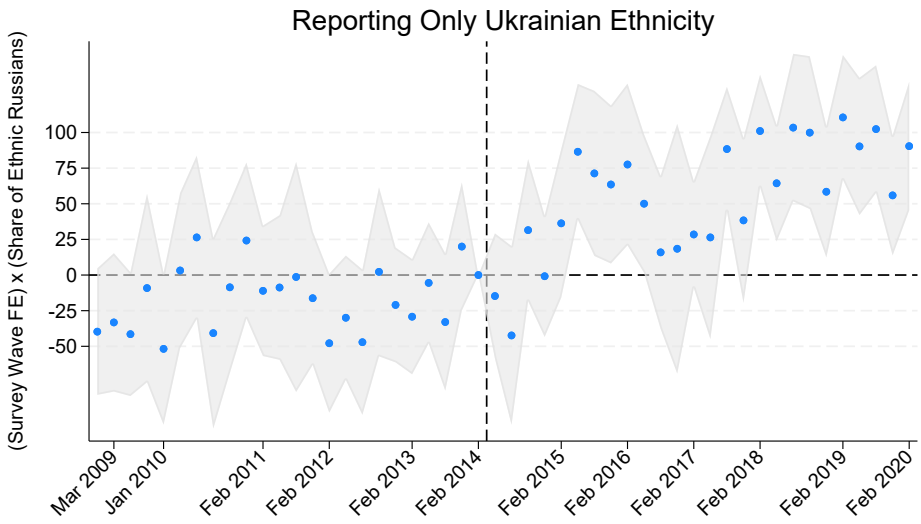


Figure 4: Event-Study by Pre-War Ethnic Composition, Ukrainian Only Ethnicity, 2014

Notes: This figure plots event-study coefficients from a regression of an indicator for reporting exclusively Ukrainian ethnicity on survey-wave indicators interacted with the pre-war share of ethnic Russians in the respondent's district, measured in the 2001 Census (Equation (2)). Positive post-2014 coefficients indicate relatively larger increases in exclusive Ukrainian identification in districts with higher pre-war shares of ethnic Russians, consistent with consolidation rather than polarization. Coefficients are normalized relative to the February 2014 survey wave, completed prior to the annexation of Crimea. The vertical line marks the onset of hostilities in late February 2014. The sample excludes Crimea and districts that were fully or partially occupied after 2014. Standard errors are clustered at the district level. Shaded areas indicate 95% confidence intervals.

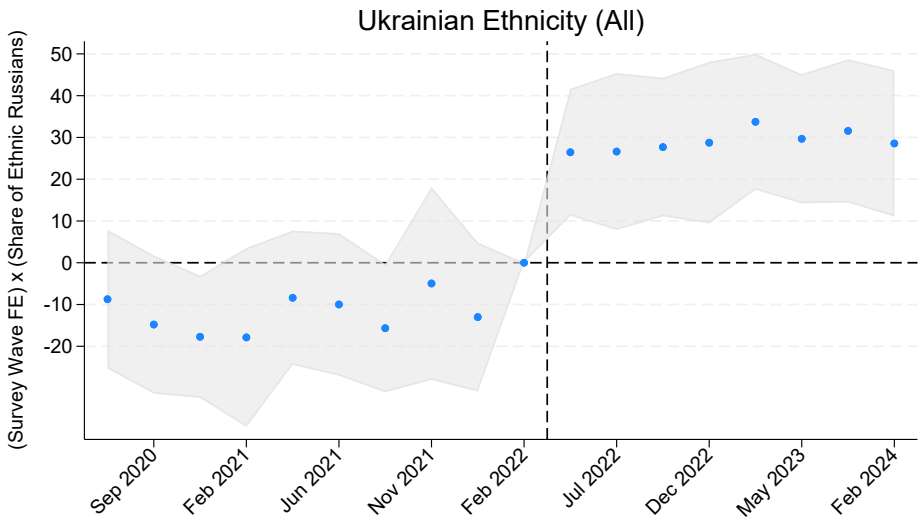


Figure 5: Event-Study by Pre-War Ethnic Composition, Ukrainian Ethnicity, 2022

Notes: This figure plots event-study coefficients from a regression of an indicator for reporting Ukrainian ethnicity on survey-wave indicators interacted with the pre-war share of ethnic Russians in the respondent's pre-2020 district, measured in the 2001 Census (Equation (2)). Positive post-February 2022 coefficients indicate relatively larger increases in exclusive Ukrainian identification in districts with higher pre-war shares of ethnic Russians, consistent with consolidation rather than polarization. Coefficients are normalized relative to the February 2022 survey wave, completed prior to the full-scale invasion. The vertical line marks the invasion onset in late February 2022. The specification includes post-2020 district fixed effects (post-2020 rayons); for internally displaced respondents, the district is defined as the pre–February 2022 district of residence. Standard errors are clustered at the post-2020 district level. Shaded areas indicate 95% confidence intervals.

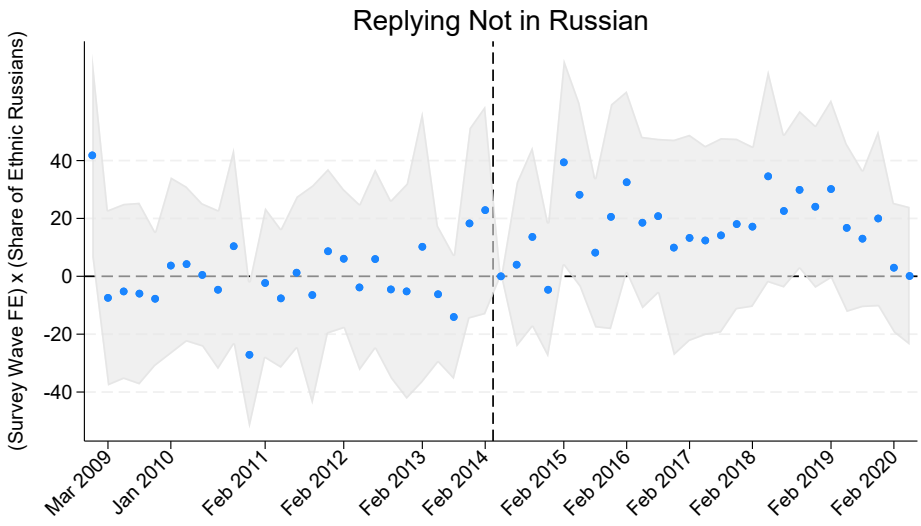


Figure 6: Event-Study by Pre-War Ethnic Composition, Replied Not in Russian, 2014

Notes: This figure plots event-study coefficients from a regression of an indicator for replying not in Russian (equal to one for replying in Ukrainian or mixed Ukrainian–Russian) on survey-wave indicators interacted with the pre-war share of ethnic Russians in the respondent's district, measured in the 2001 Census (Equation (2)). Positive post-2014 coefficients indicate relatively larger increases in non-Russian replies in districts with higher pre-war shares of ethnic Russians, consistent with consolidation rather than polarization. Coefficients are normalized relative to the February 2014 survey wave, completed prior to the annexation of Crimea. The vertical line marks the onset of hostilities in late February 2014. The sample excludes Crimea and districts that were fully or partially occupied after 2014. Standard errors are clustered at the district level. Shaded areas indicate 95% confidence intervals.



Figure 7: Event-Study by Pre-War Ethnic Composition, Replied Not in Russian, 2022

Notes: This figure plots event-study coefficients from a regression of an indicator for replying not in Russian (equal to one for replying in Ukrainian or mixed Ukrainian–Russian) on survey-wave indicators interacted with the pre-war share of ethnic Russians in the respondent's pre-2020 district, measured in the 2001 Census (Equation (**??**)). Positive post-February 2022 coefficients indicate relatively larger increases in non-Russian replies in districts with higher pre-war shares of ethnic Russians, consistent with consolidation rather than polarization. Coefficients are normalized relative to the February 2022 survey wave, completed prior to the full-scale invasion. The vertical line marks the invasion onset in late February 2022. The specification includes post-2020 district fixed effects (post-2020 rayons); for internally displaced respondents, the district is defined as the pre–February 2022 district of residence. Standard errors are clustered at the post-2020 district level. Shaded areas indicate 95% confidence intervals.

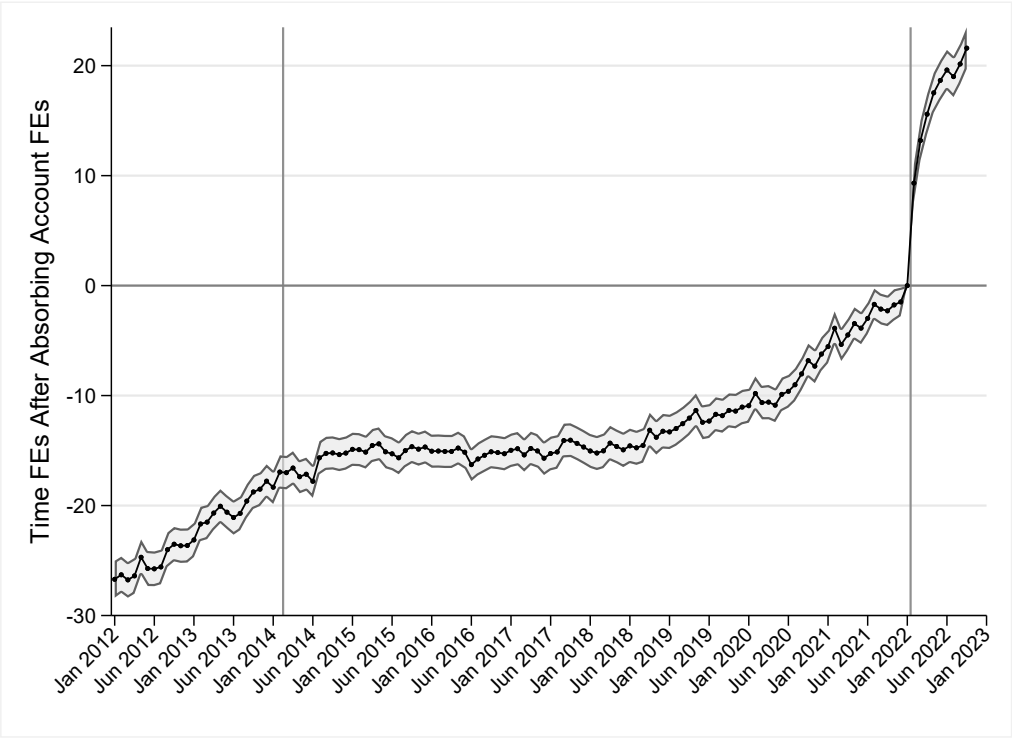


Figure 8: Within-Account Shift in Use of Ukrainian Language on Twitter Over Time

Notes: The figure plots estimated month fixed effects from a panel of 36 million tweets geolocated to Ukraine (January 2012–November 2022). Language is assigned using Twitter's built-in detector. Restricting attention to tweets labeled Ukrainian or Russian, we compute for each account the monthly share of Ukrainian-language tweets and estimate a regression with account fixed effects and month indicators. The coefficients shown are the estimated month effects, normalized to January 2022 (the last pre-invasion month), and capture within-account changes in the propensity to tweet in Ukrainian, net of time-invariant differences across accounts. Vertical lines mark February 2014 and February 2022. Shaded areas indicate 95% confidence intervals.

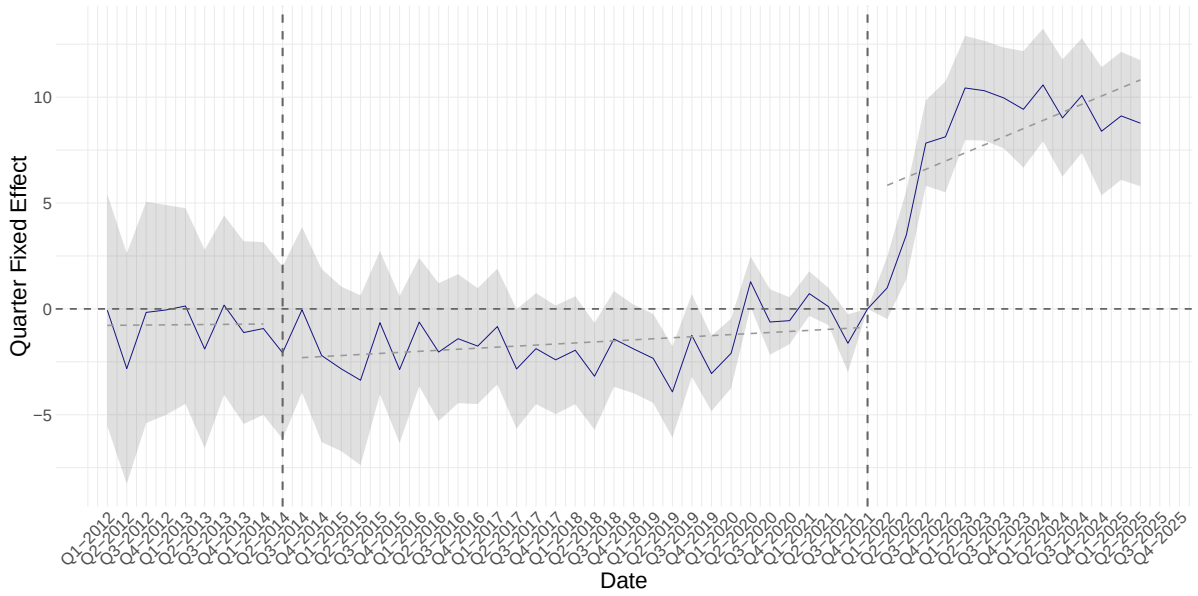


Figure 9: Shift toward Ukrainian in Google Search Language Over Time

Notes: The figure plots year–quarter fixed effects for the share of Google searches conducted in Ukrainian for a given word in a given region and quarter, normalized to 2021Q4 (the last pre-invasion quarter). To flexibly account for seasonality, the specification includes calendar-quarter-by-word-by-region fixed effects. As a result, the reported coefficients are identified from within-year deviations relative to each word–region pair's recurring seasonal pattern. The regression additionally controls for word- and region-specific linear trends estimated over the pre-2022 period. The resulting coefficients therefore represent deviations from both seasonal patterns and pre-invasion trends. The sample is restricted to regions under Ukrainian government control. Observations are weighted by word–region search volume. Standard errors are clustered at the word–region level; shaded areas denote 95% confidence intervals.

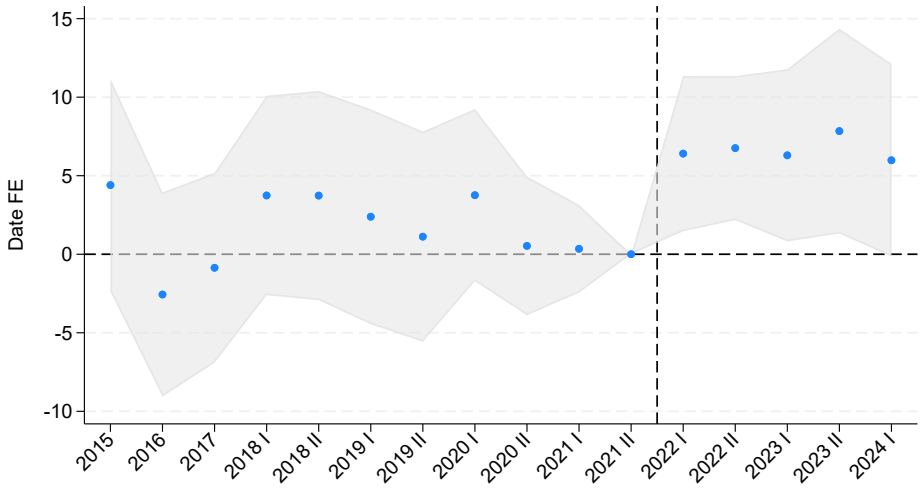


Figure 10: Time-Varying Association Between Name Ukrainianness and Popularity

Notes: The figure plots event-study coefficients from Equation (3), which estimates how the association between a name's pre-war Ukrainianness and its popularity among newborns evolves over time. The unit of observation is name $n \times$ region $r \times$ period t (half-year or year). The dependent variable, $Popular_{rtn}$, is an indicator equal to one if name n appears in the Ministry of Justice list of most popular names in region r and period t. The specification includes region–period and region–name fixed effects. Coefficients correspond to interactions between period indicators and a name-level Ukrainianness index and are normalized to 2021 II (the last pre-invasion period). Positive coefficients indicate that names with higher Ukrainianness are relatively more likely to appear among popular names in that period. The sample excludes regions occupied in 2022. Standard errors are clustered at the region level; shaded areas denote 95% confidence intervals.

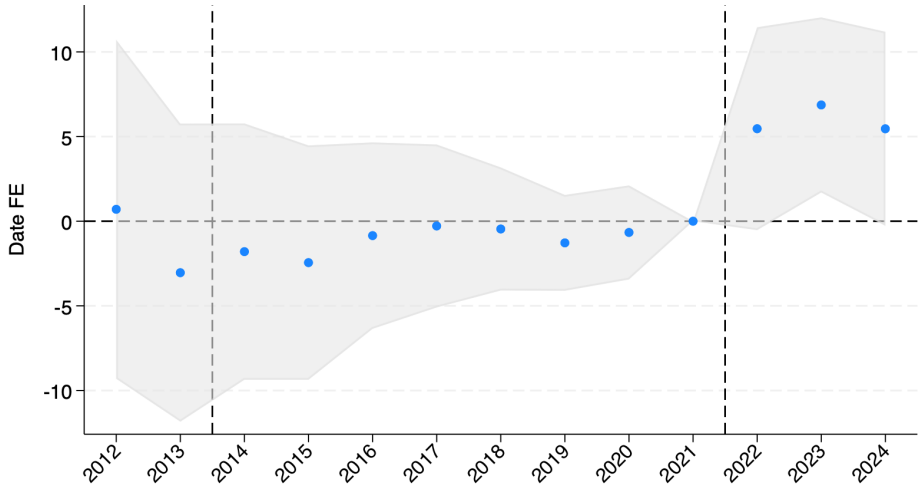


Figure 11: Time-Varying Association Between Dish Ukrainianness and Search Intensity

Notes: The figure plots yearly event-study coefficients from Equation (4), which estimates how the association between a dish's pre-war Ukrainianness and its search intensity evolves over time. The outcome is the raw Google Trends index for dish i in oblast r and month t, aggregated to the yearly level. The specification includes dish-by-oblast and oblast-by-month fixed effects. Coefficients are normalized to 2021, the omitted reference year, so estimates represent deviations from the pre-invasion Ukrainianness gradient in search intensity. Vertical lines mark the onset of hostilities in 2014 and the full-scale invasion in 2022. Standard errors are two-way clustered by dish and oblast; shaded areas denote 95% confidence intervals.

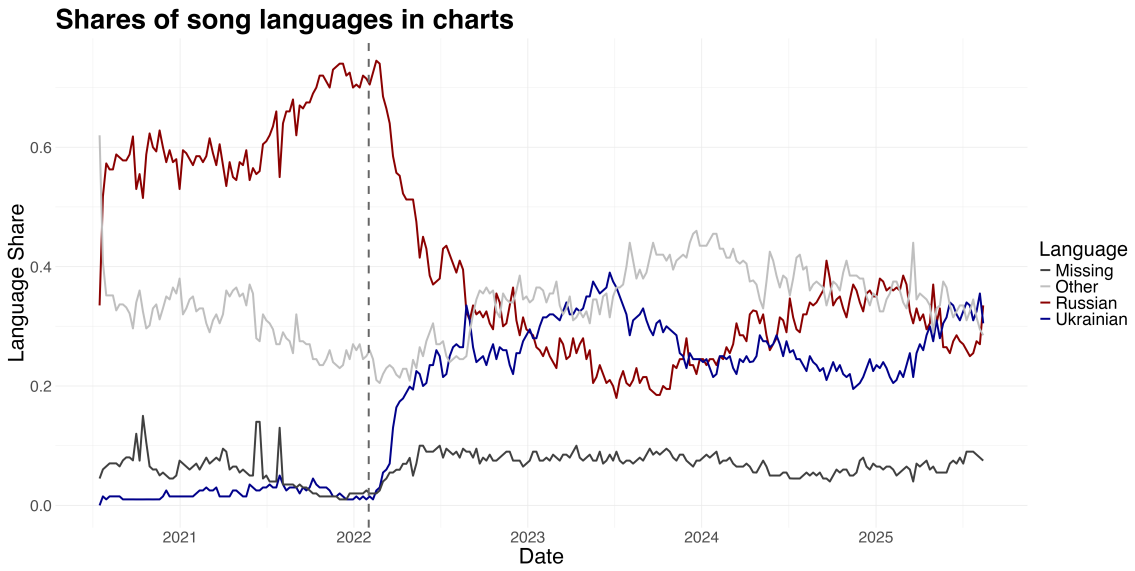


Figure 12: Spotify Weekly Top Songs Ukraine: Language Composition of Chart Entries, 2020–2025

*Notes:*The figure reports, at the weekly frequency, the share of tracks in the Spotify Weekly Top Songs Ukraine chart by language (Ukrainian, Russian, or other). Track language is determined using lyric-based language detection from Genius; tracks without available lyrics (e.g., instrumentals or songs not documented on Genius) are classified separately when relevant.

A Literature Review

A.1 Research in Sociology, Political Science, and Psychology

The relationship between nation-building and military conflict has received sustained attention across the social sciences. Seminal work in sociology and political science emphasizes the role of war in state formation and the construction of national identity. In modernist and statist accounts, large-scale military conflict raises states' demands for taxation, conscription, and coordination, prompting elites to invest in public goods and promote a shared national identity to sustain mass mobilization (Tilly, 1975; Mann, 1993). Related work highlights how modernization and industrialization generate pressures for cultural standardization, leading to the construction of common languages and "high cultures" that facilitate national integration (Gellner, 1983; Anderson, 1983). In these perspectives, nationalism is largely a product of modernity, shaped by elite responses to the demands of war and governance (Smith, 1998; Hutchinson, 2017).

In contrast, ethnosymbolist approaches emphasize continuity rather than invention. Smith (1986) argues that modern nations draw on pre-modern myths, symbols, and collective memories, and that shared rituals of commemoration provide the foundations for durable national identities. Under this view, the shared trauma and affective potency associated with wars often consolidate national identities through bottom-up processes rooted in collective experience rather than elite instrumentalization (Smith, 1981; Hutchinson, 2017).

At the same time, a substantial body of comparative and case-based research documents that conflict can weaken national identification when violence overlaps with internal ethnic, linguistic, or regional cleavages. Wimmer (2012) traces the global emergence of nation-states, providing empirical support for statist theories, while Darden and Mylonas (2016) highlight the role of external military threats in expanding national schooling. Yet other work shows that conflict can reinforce subnational identities. For example, Hughes (2011) demonstrates how British state violence during

the Troubles entrenched Irish nationalism and radicalized communities. Similarly, Sambanis and Shayo (2013) find that civil conflict can weaken national attachment in favor of increased ethnic identification. Relatedly, Sambanis et al. (2015) argue that victory in interstate conflict strengthens national identity, while defeat erodes it.

Micro-level research further clarifies the mechanisms underlying these divergent outcomes. Brewer (1999) distinguishes between ingroup attachment and outgroup hostility, showing that national identification need not entail animus toward minorities, although Rui J. P. de Figueiredo and Elkins (2003) find that patriotic pride often correlates with prejudice. Fritsche et al. (2017) show that traumatic events increase reliance on ingroups by heightening feelings of personal helplessness. Experimental evidence further suggests that external threats can raise national identification primarily by increasing hostility toward outgroups rather than by deepening attachment to the nation itself (Sambanis and Lee, 2023).

Taken together, this literature establishes that the relationship between violent conflict and national identity is theoretically ambiguous and empirically context-dependent. While some forms of war foster national consolidation, others intensify internal divisions. Crucially, existing studies rarely distinguish consolidation from polarization within the same country or trace identity change across both attitudes and behavior, motivating the empirical approach adopted in this paper.

A.2 Economic Research

(a) General Approaches. Alesina et al. (2020) formalize the nation-building role of wars emphasized by Tilly and Mann; they argue that governments at war expand public goods and invest in nation-building to rally citizens, making identity formation a strategic policy choice. Relatedly, Aghion et al. (2019) document that investments in mass primary education rise with external military rivalry, and that this response is stronger in democracies. Complementing this external-threat channel, Alesina et al. (2021) study when rulers deliberately homogenize the population and argue that the strongest incentives arise when elites are threatened by democratization; they empha-

size mass primary education as a central instrument of nation-building. Relatedly, Almagro and Andrés-Cerezo (2020) show theoretically that nation-building's success at supplanting regional identities is contingent on economic, social, and political factors.

In contrast, Shayo (2009) models identity formation as an endogenous choice between class and nation. Individuals balance similarity to other group members with the relative status of each group, yielding equilibria in which the poor may either identify with their class and support high redistribution or with the nation and accept lower redistribution. National eminence, external threats, or heterogeneity within the lower class can shift identification toward the nation, dampening redistributive demands. This framework is especially relevant for studying Ukraine, where conflict and external threat heighten the salience of national identity and reshape the trade-offs between alternative self-conceptions.

(b) Comparative and Individual Case Studies. Previous empirical studies in economics have examined the role of language and the tension between national and narrower (ethnic or regional) identities. With regards to the former, Fouka (2020) demonstrates that post-WWI language bans targeting ethnic Germans in the United States produced backlash rather than assimilation, while Clots-Figueras and Masella (2013) find that bilingual education in Catalonia reinforced a regional identity, even among ethnic Spaniards. A complementary approach emphasizes intergroup contact rather than direct language policy: exploiting quasi-random exposure to Indonesia's Transmigration program, Bazzi et al. (2019) show that policy-induced local diversity can foster long-run integration, including greater use of the national language and other markers of assimilation. With regards to the latter, Depetris-Chauvin et al. (2020) demonstrate that football victories in Sub-Saharan Africa raise national identification and reduce interethnic violence, whereas Dehdari and Gehring (2022) show that the repressive nation-building efforts in Alsace-Lorraine (by the German Empire after 1871 and by the French Republic after 1918) reinforced and entrenched regional identities.

Beyond nation-building policy, Jha (2013) shows that interethnic economic complementarities can sustain long-run coexistence: South Asian medieval trading ports, where Hindu-Muslim exchange was institutionalized, were substantially less prone to Hindu-Muslim riots long after the original trade advantages dissipated.

DellaVigna et al. (2014) specifically focus on the (legacy of) military conflict, finding that access to cross-border Serbian radio in Croatia increased support for extreme nationalist parties and ethnically offensive expression. This result indicates that conflict can reinforce national identification while also bolstering outgroup sentiment. Relatedly, Cannella et al. (2024) examines the German temporary annexation of parts of Northern Italy and finds persistent increases in political polarization and electoral support for both left- and right-wing extremist parties.

A distinct line of work studies large-scale state-building interventions in diverse societies. Carlitz et al. (2025) analyze Tanzania's Ujamaa villagization and political education campaign and find persistent increases in national identity and perceived state legitimacy among exposed cohorts, alongside trade-offs: exposure lowers demands for democratic accountability and does not increase generalized interethnic trust. In a related spirit, Bazzi et al. (2025) study Indonesia's mass public school construction program and show that state expansion induced competitive entry by Islamic schools rather than crowd-out, with long-run consequences for political and cultural outcomes, including higher religiosity.

Finally, Ananyev and Poyker (2023) highlight why the relationship between conflict and national identity is theoretically ambiguous. Studying the Tuareg rebellion in Mali, they show that proximity to conflict reduced identification with the nation and strengthened ethnic identities among groups not directly engaged in the fighting. Their findings suggest that conflict can undermine rather than consolidate the nation, particularly when it exposes the weakness of state institutions. This insight underscores a key point for our study: the direction of identity change in response to violent conflict cannot be determined *ex ante*, making Ukraine a crucial case for understanding the conditions under which war either consolidates or fragments national identity.

(c) Agent-Level. Microeconomic theory models identity as both a preference and a constraint. Individuals derive utility from conforming to group norms and from the esteem associated with belonging, while deviation produces disutility (Akerlof and Kranton, 2000; Kranton, 2016). Because identities are partly chosen and partly imposed, people invest in self-conceptions that reinforce their place within valued groups, particularly when those identities are contested or threatened (Bénabou and Tirole, 2011). Utility itself depends on the standing of the group, so that people often optimize as members of a collective rather than as isolated individuals (Akerlof, 2016). This rationale clarifies why external threats are powerful triggers of nation-building. When conflict raises the stakes of group membership, individuals are more likely to adopt the behaviors that signal alignment with the nation.

Empirical work provides additional insights. Bauer et al. (2016) synthesize evidence from surveys and behavioral experiments across a wide range of settings and conclude that exposure to wartime violence tends to increase local cooperation and civic engagement, with suggestive evidence that these effects are often parochial rather than universal. Weisel and Zultan (2016) show experimentally that when individuals perceive their group as the target of threat, they reallocate effort and loyalty toward the ingroup. Yet Gehring (2021) demonstrates that Russian aggression in 2014 increased European identification in areas closer to Russia, illustrating how external threats can reorient supranational instead of purely national attachments.

A.3 Measuring National Identity

Recent work studying individuals' national identities has introduced a host of measurement approaches complementing traditional surveys. Social media offers a source of high-frequency data, with shifts in language use and online self-presentation serving as barometers of national identity (Abramenko et al., 2024). Metzger et al. (2016) follow a similar approach by analyzing Twitter language choice during the Euromaidan protests, treating users' preferred and actual posting languages as proxies for ethnolinguistic identification. Naming choices serve as a complementary, durable

measure. Since names are public and costly to reverse, changes in their ethnic distinctiveness capture individuals' national identities (Fouka, 2019; Esposito et al., 2023; Assouad, 2025). Consumption behavior serves as a third outcome: Atkin et al. (2021) infer ethnic and religious identity from food consumption patterns in India, while Nardotto and Sequeira (2025) exploit the Brexit referendum to show that shifts in the salience of national identity affect consumers' relative demand for domestic versus foreign goods. Recent work has shown that conflict-related shocks during the Iraq War exerted changes on food brand consumption in the US (Pandya and Venkatesan, 2016; Helms et al., 2025). Our paper relies on, and develops, all three of these measurement approaches.

B Additional Details on Data Construction

B.1 KIIS Survey Data and Variable Definitions

Our baseline analysis uses repeated cross-sections from the Kyiv International Institute of Sociology (KIIS) omnibus surveys, fielded quarterly between 2008 and 2024. The data comprise 72 survey waves and more than 156,000 respondents and are designed to be nationally representative of the adult population. We restrict the sample to respondents residing in territories under Ukrainian government control at the time of each survey wave.[17]

B.1.1 Survey-Based Identity Outcomes.

We study three survey-based identity outcomes that are consistently available over time: two measures of self-reported ethnicity ("nationality") and the language used during the interview.

Nationality (open-ended). KIIS first asks respondents an open-ended question on self-identified nationality:

> ***Question:*** *"What is your nationality, as you consider yourself?"*

Responses are subsequently coded by KIIS into predefined categories, including Ukrainian, Russian, both Ukrainian and Russian (only when explicitly stated by the respondent), Belarusian, Jewish, other, and no reply.

Nationality gradient. For respondents identifying as Ukrainian and/or Russian, KIIS additionally administers a more granular ordered question (available through early 2022) intended to capture mixed and intermediate identities:

> ***Question:*** *"I will ask you to answer one more question about nationality, in order to record it as accurately as possible. This is because some people consider*

[17] For an example of KIIS wave documentation and sampling procedures, see the June 2021 omnibus survey materials.

themselves to belong to two nationalities at the same time. Please look at this card and tell me which of these statements best corresponds to your opinion of yourself. I consider myself…"

The response categories are: (i) Only Ukrainian; (ii) Both Ukrainian and Russian, but more Ukrainian; (iii) Equally Ukrainian and Russian; (iv) Both Russian and Ukrainian, but more Russian; (v) Only Russian.

Interview language. Finally, KIIS records the language used during the interview following a standardized protocol:

> **Instructions:** Interviewers greet respondents neutrally, ask which language is easier for them, and conduct the interview in the chosen language.

Interview language is coded as Ukrainian, Russian, or mixed Ukrainian–Russian. Because this measure reflects a revealed linguistic choice in a low-stakes interaction rather than a direct attitudinal response, it provides a useful complement to self-reported identity measures.

In addition to identity outcomes, we also use a measure of attitudes toward ethnic minorities based on a standard "social distance" battery (e.g., willingness to accept members of a minority group as family members, close friends, neighbors, coworkers, residents or guests of Ukraine, or not admit them to Ukraine). This outcome is coded as an ordered index and, in robustness checks, as a set of category indicators.

B.2 Google Trends Data

B.2.1 Language Use in Google Searches

Word List Construction. To construct comparable Ukrainian–Russian search measures, we begin by downloading text corpora from the Leipzig Corpora Collection (Goldhahn et al., 2012; Leipzig Corpora Collection, 2025). For Ukrainian, we use the 2019 corpus consisting of one million sentences drawn from Ukrainian websites; for Russian, we use the 2015 corpus of one million sentences collected from websites in Ukraine. From each corpus, we extract nouns only.

Each Ukrainian noun is translated into Russian, and each Russian noun is translated into Ukrainian, producing two translation lists. We then match translations across lists to form pairs of words with identical meanings but different spellings. Words that are spelled identically in both languages are excluded. We retain the 198 most frequently used unique word pairs for analysis.

Google Trends Queries and Aggregation. For each word pair i, region r, and month t, we retrieve Google Trends indices for both the Ukrainian and Russian spellings over the period 2012–June 2025. We aggregate monthly data to the quarterly level to reduce noise. Let $popularity_{tir}^{uk}$ and $popularity_{tir}^{ru}$ denote the corresponding search indices.

Outcome Construction. We define the relative share of Ukrainian spelling as:

$$S_{tir} = 100 \times \frac{popularity_{tir}^{uk}}{popularity_{tir}^{uk} + popularity_{tir}^{ru} + \epsilon}, \quad \epsilon = 10^{-3}$$

We additionally construct weights reflecting the relative importance of each word pair in total search activity within a region:

$$w_{i,r} = \frac{\sum_t \left(popularity_{tir}^{uk} + popularity_{tir}^{ru}\right)}{\sum_{ti} \left(popularity_{tir}^{uk} + popularity_{tir}^{ru}\right)}.$$

Our baseline specifications weight observations by w_{ir}, though results are robust to unweighted alternatives.

B.2.2 Interest in National Cuisine

Dish Selection. We measure cultural consumption using Google search interest in national cuisine. The dish set comprises 100 items compiled from Wikipedia lists of Ukrainian and Russian cuisine.

Query Construction. For each dish, we retrieve Google Trends indices by oblast from 2012 to 2024 at monthly frequency. To avoid mechanically loading cuisine on

language choice, each query combines Ukrainian- and Russian-language spellings using an OR operator, yielding a single popularity series per dish.

Ukrainianness Index. Since Google Trends normalizes search intensity within each query, indices are not directly comparable across independent searches. Because only a limited number of terms can be queried jointly, we use Borsch as a fixed reference term and scale all other dishes relative to it. This procedure allows us to construct consistent relative popularity distributions across countries and over time. Each dish i is assigned a time-invariant Ukrainianness score:

$$U_i \equiv \frac{\bar{s}_i^{UKR}}{\bar{s}_i^{UKR} + \bar{s}_i^{RUS}},$$

where $\bar{s}_i^c$ denotes the dish's average pre-war share of search intensity within country c, computed by normalizing dish indices to sum to one within each country-month and averaging over 2012–2021. Higher values of U_i indicate dishes searched disproportionately more in Ukraine than in Russia prior to the war.

Validation. We assess the external validity of the cuisine Ukrainianness index using within-Ukraine cross-sectional variation. Appendix Figure B.1 plots, for each oblast, the average Ukrainianness of dishes implied by pre-war search behavior against the oblast's share of ethnic Russians in the 2001 Census. The relationship is strongly negative: oblasts with a larger pre-war ethnic Russian presence display systematically lower baseline Ukrainianness in cuisine-related searches. This pattern is consistent with the intended interpretation of the index and mirrors the validation exercise for baby names. Occupied regions (Crimea, Donetsk, Luhansk, and Sevastopol) are excluded.

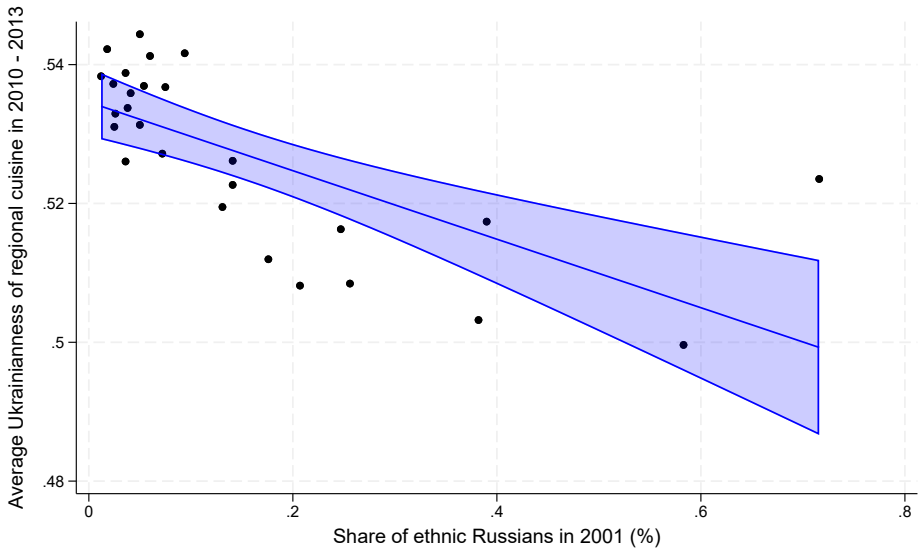


Figure B.1: Validating the Cuisine Ukrainianness Index within Ukraine

Notes: This figure assesses the external validity of the cuisine *Ukrainianness* index within Ukraine. Each point is a region (oblast) and plots the oblast's average pre-war cuisine Ukrainianness—computed as the search-share-weighted mean of dish-level Ukrainianness using Google Trends data from 2010–2013—against the oblast's share of ethnic Russians in 2001 (Census). The solid line is the linear fit; the shaded area is the 95% confidence band. Consistent with the interpretation of the index, regions with a higher pre-war share of ethnic Russians tend to display lower Ukrainianness of cuisine-related searches.

Outcomes. Let Pop_{irt}^{raw} denote the Google Trends index for dish i in oblast r and month t. We also report a standardized outcome:

$$Pop_{irt}^{std} = \frac{Pop_{irt}^{raw} - \overline{Pop_{i,2012–2021}^{raw}}}{sd(Pop_{i,2012–2021}^{raw})},$$

computed dish-by-dish using the pre-war distribution.

Dish-level gradient. Finally, the dish-level classification plot in Figure B.2 summarizes the cross-dish relationship between war-period changes in search intensity and U_i. In a first step, we estimate dish-specific post-2022 changes in search intensity in a dish-region-month panel,

$$\text{Popularity}_{irt} = \alpha_{ir} + \gamma_{rt} + \delta_i \cdot \text{Post}_t + \varepsilon_{irt},$$

where α_{ir} are dish-region fixed effects and γ_{rt} are region-time fixed effects. The coefficients δ_i capture dish-specific deviations in the post period relative to the pre-period, net of region-specific time shocks. In a second step, we relate these estimated dish-level shifts to the Ukrainianness score,

$$\delta_i = \beta U_i + \eta_i,$$

so that β captures whether more-Ukrainian dishes experienced systematically larger increases in search intensity after 2022.

Within the dish set, the Ukrainianness index aligns with common culinary associations: canonical Ukrainian dishes such as borscht and holubtsi receive high values, while dishes more closely associated with Russian cuisine receive low values. Figure B.2 shows a clear positive gradient in the post-2022 period: estimated changes in search intensity δ_i are increasing in U_i. This pattern implies a systematic reallocation of attention toward more Ukrainian dishes, rather than changes concentrated in a small number of queries.

B.2.3 Baby names: data construction and validation

Ministry of Justice lists. We collect lists of the most popular baby names by region and half-year from the Ukrainian Ministry of Justice for 2015–2024. The number of names listed varies across region–period cells. The resulting panel has three dimensions (name n, region r, period t).

Name Ukrainianness. Each name is assigned a time-invariant Ukrainianness score based on its pre-war relative popularity in Ukraine versus Russia using *Forebears.io*:

$$\text{Ukrainianness}_n = \frac{U_n}{U_n + R_n},$$

where U_n and R_n are the pre-war shares of name n in Ukraine and Russia, respectively. We treat alternative spellings/transliterations of the same name as equivalent (e.g.,

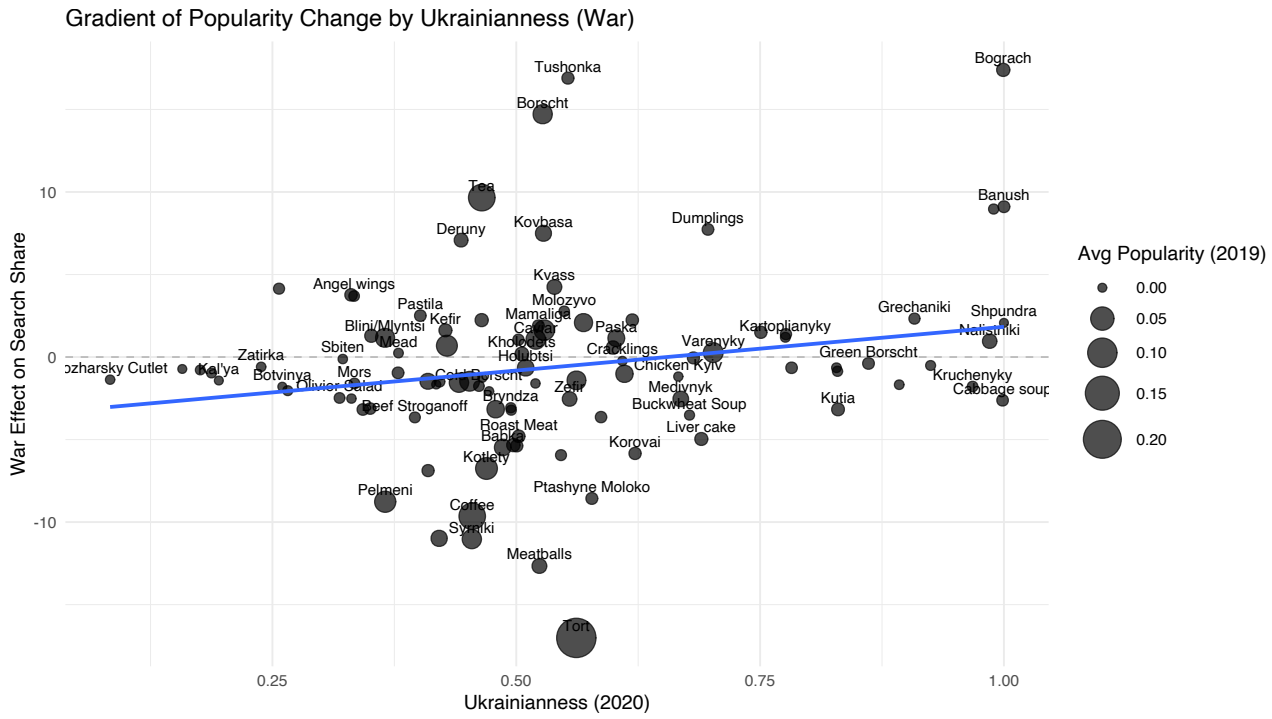


Figure B.2: 2022 War Change in Dish Search Share versus Dish Ukrainianness.

Notes: Each point represents a dish. The horizontal axis reports the dish's Ukrainianness score, constructed from pre-war relative search shares in Ukraine and Russia. The vertical axis plots the dish-level change in search intensity during the war period, measured as the change in the dish's share of total searches within the 100-dish set between the pre-war baseline (2012–2021) and the war period (2022–2024), using the same Google Trends measure as in the main analysis. Marker style indicates dish classifications (e.g., prepared dishes vs. broad categories vs. spike-driven items). The fitted lines summarize the cross-dish relationship between war-period changes in search intensity and Ukrainianness score within each classification; a positive slope indicates larger increases for more-Ukrainian dishes.

"Serhii" and "Sergey").

The dataset contains 211 unique names (100 boys, 111 girls). The Ukrainianness index is approximately uniformly distributed on $[0, 1]$; its first quartile, median, and third quartile are 0.21, 0.50, and 0.84.

Validation. Appendix Figure B.3 plots each region's average Ukrainianness (computed over 2015–2021) against the region's share of ethnic Russians in the 2001 Census. The fitted relationship is strongly negative: regions with a larger pre-war Russian ethnic presence exhibit lower baseline Ukrainianness of popular baby names, consistent with the intended interpretation of the measure. Crimea, Donetsk, Luhansk, and

Sevastopol are excluded from this validation exercise.

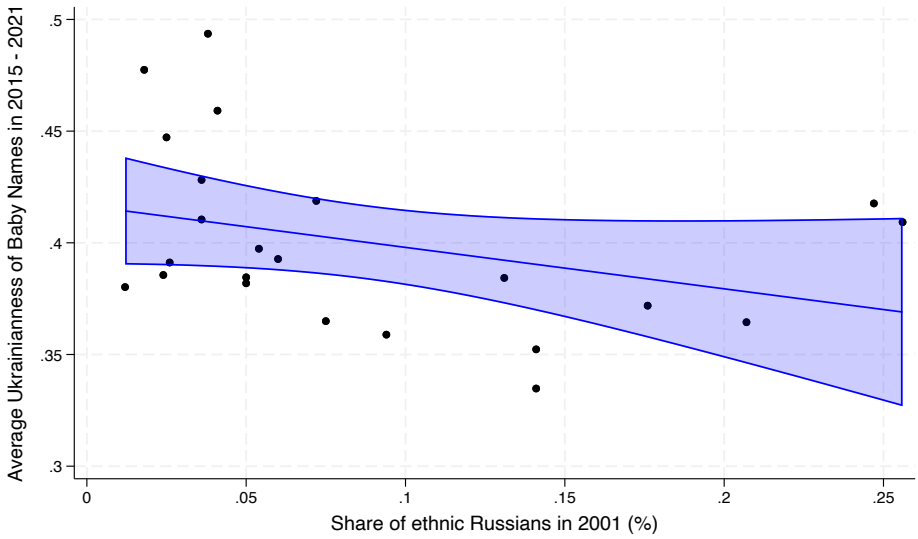


Figure B.3: Validating the baby-name Ukrainianness index within Ukraine

Notes: This figure assesses the external validity of the baby-name *Ukrainianness* index within Ukraine. Each point is an region (oblast) and plots the oblast's average Ukrainianness of names appearing in the Ministry of Justice popularity lists over 2015–2021 against the oblast's share of ethnic Russians in 2001 (Census). Occupied at least partially regions (Crimea, Donetsk, Luhansk, and Sevastopol) are excluded. The solid line is the linear fit; the shaded area is the 95% confidence band. Consistent with the interpretation of the index, regions with a higher pre-war share of ethnic Russians tend to have lower average Ukrainianness of popular baby names.

Alternative specification. As a more transparent alternative to the main specification, we estimate half-year effects for the average Ukrainianness of popular names at the region×period level (restricting to regions not occupied in 2022):

$$\text{Ukrainianness}_{rt} = \sum_{s \neq 2021\ \text{II}} \beta_s \mathbb{I}[\text{period} = s] + X'_{rt}\Gamma + \alpha_r + \varepsilon_{rt}, \tag{6}$$

where the second half of 2021 is the omitted period and α_r are region fixed effects. Controls $X_{r,t}$ include the number of distinct names listed in region r and period t (allowing for a region-specific slope δ_r), as well as region-by-period measures of internal displacement inflows and a proxy for refugee outflows. As one may notice this specification allows us to explicitly control for population displacement. Standard errors are clustered by region.

Figure B.4 exhibits a classic event-study pattern: coefficients are near zero and mostly statistically indistinguishable from the zero, followed by a sharp and sustained increase beginning in 2022 I. This indicates that, controlling for region fixed effects, cell size, and refugee/IDP exposure, the average Ukrainianness of popular names rose markedly after the full-scale invasion of February 24, 2022.

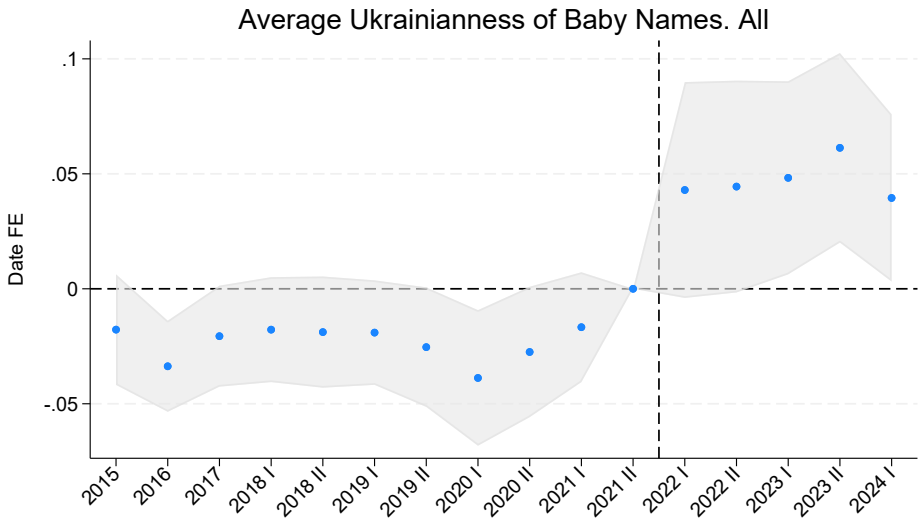


Figure B.4: Average Ukrainianness of popular baby names over time (non-occupied regions)

Notes: The figure plots half-year coefficients from equation (??), which estimates changes in the average Ukrainianness of popular baby names across regions over time. The outcome, $\text{Ukrainianness}_{r,t}$, is the mean Ukrainianness index of names appearing in the Ministry of Justice popular-name list in region r and period t. The specification includes region fixed effects, region-specific slopes in the number of listed names ($N_{r,t}$), and controls for refugee inflows and the share of internally displaced persons. Coefficients are normalized to zero in 2021 II (the omitted reference period) and therefore represent deviations relative to the pre-invasion period. The sample is restricted to regions not occupied in 2022. Standard errors are clustered at the region level; whiskers denote 95% confidence intervals.

Name-level gradient. To assess whether the effect is broad-based across names, we estimate name-specific post-war changes in popularity after absorbing region–time and region–name fixed effects:

$$\text{Popular}_{rtn} = \tau_n \mathbb{I}[t \geq \text{Feb 2022}] + \alpha_{rt} + \eta_{rn} + u_{rtn}, \tag{7}$$

and relate the estimated $\hat{\tau}_n$ to Ukrainianness$_n$ (unweighted and weighted by pre-2022 popularity):

$$\hat{\tau}_n = \alpha + \beta \, \text{Ukrainianness}_n + \varepsilon_n. \tag{8}$$

Figure B.5 shows a positive and statistically significant gradient, indicating that the post-2022 reallocation toward Ukrainian names is not driven by a handful of outliers.

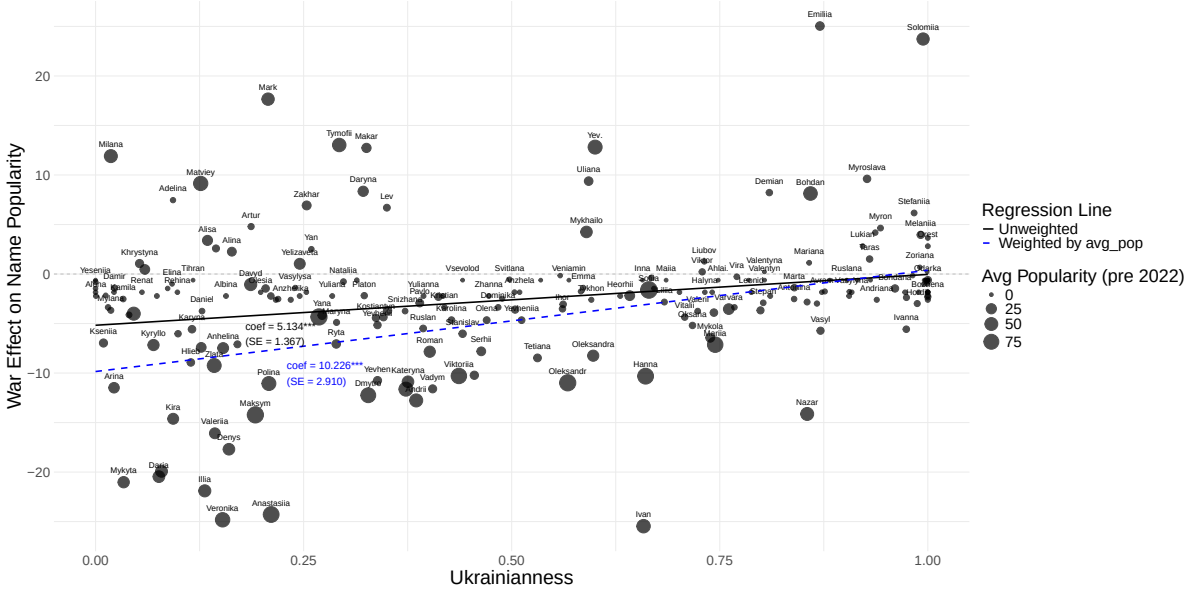


Figure B.5: Name-Level War Effects and the Ukrainianness Correlation

Notes: Each point represents a name. The vertical axis plots the estimated name-specific post-February 2022 effect $\hat{\tau}_n$ from equation (7), which is obtained from a difference-in-differences specification with region–time and region–name fixed effects. The horizontal axis reports the name's pre-war Ukrainianness score. Solid lines show the cross-sectional gradient from equation (8), estimated (i) unweighted and (ii) weighted by the name's average pre-2022 popularity. Marker size is proportional to pre-2022 average popularity. A positive slope indicates that more-Ukrainian names experienced larger post-2022 increases in popularity.

B.2.4 Music Consumption: Spotify Charts and Artist-Level Data

Scope and Main Sources. We measure music consumption using Spotify Weekly Top Songs charts for Ukraine (Spotify, 2025a). Historical weekly chart data are scraped from Kworb (2025). The sample covers July 17, 2020 (Spotify's official entry into the Ukrainian market) through August 14, 2025. To enrich chart entries with song metadata and release dates, we use the Genius API (Genius, 2025) and the Spotify

Web API (Spotify, 2025c). For the supply-side analysis, we additionally construct a list of Ukrainian artists using NUAM Club (2025), Spotify (2025b), and Wikipedia contributors (2025).

Chart Panel Construction. The raw weekly chart data contain song titles, artist names, chart dates, chart positions, and weekly stream counts for songs ranked in Spotify's Top-200 chart in Ukraine. We reshape these weekly snapshots into a song-week panel. The final chart dataset contains 3,850 distinct tracks performed by 572 artists. The baseline unit of observation is a song-week chart entry.

Song Matching and Metadata Enrichment. We harmonize song titles and artist names across sources and construct song-level identifiers used to merge metadata. We then retrieve release dates and additional song metadata from Genius and the Spotify Web API. These release dates are used to define pre-war and post-war release subsamples (e.g., restricting to songs released before 2022 to isolate demand-side shifts in listening behavior).

Lyrics Collection and Language Classification. We classify charting songs by language using lyrics retrieved via the Genius API (Genius, 2025). After matching each charting track to a Genius entry, we apply automated language detection to the retrieved lyrics. For the main analysis, songs are assigned to mutually exclusive language groups: *Ukrainian*, *Russian*, *Other*, and *Missing lyrics*. The *Other* category includes primarily English-language songs and other non-Ukrainian/non-Russian tracks. Songs with unavailable lyrics are retained in the dataset and tracked separately.

Popularity Measures. Using the song-week panel, we construct week-by-language measures of chart presence and popularity. The main outcomes are: (i) the share of Top-200 entries belonging to each language group, and (ii) the total number of weekly streams aggregated within each language group. These measures are computed analogously for subsamples (e.g., pre-2022 releases only; Ukrainian artists only).

Pre-War Release Restriction (Demand-Side Sample). To reduce the influence of post-invasion changes in the composition of newly released songs, we construct a restricted chart sample that excludes songs released in 2022 or later. This pre-2022 release sample is used to isolate demand-side changes in listening behavior while holding the set of potentially consumed songs closer to the pre-war supply.

Artist Classification by Attitude Toward the War. For the analysis by artists' political stance, we classify artists based on Stars About War Project (2025). We match artists appearing in the Ukrainian Spotify charts to this source and assign available classifications (e.g., pro-war, anti-war) at the artist level. The resulting artist-level classification is merged onto the song-week panel through chart artist identifiers. Artists not matched to the source remain unclassified in this exercise.

Ukrainian Artist List (Supply-Side Sample). To study supply-side language choice, we construct a list of artists of Ukrainian origin whose songs appear in the Spotify Ukraine charts. We begin from the union of artist names collected from NUAM Club (2025), Spotify (2025b), and Wikipedia contributors (2025), then harmonize names and restrict to artists with at least one chart appearance in our sample. This procedure yields a final list of 266 Ukrainian artists.

Artist Catalogue Data and Release-Level Panel. For each artist in the Ukrainian-artist list, we retrieve the available catalogue from Genius (via Genius, 2025), including song titles, lyrics (when available), and release dates. We apply the same lyrics-based language classification procedure used for charting songs. This produces a release-level dataset with artist identifiers, release dates, and language labels, which is used to study changes in the language composition of newly released music.

Artist-Quarter Panel for Within-Artist Language Choice. We aggregate the release-level dataset to an artist-by-quarter panel. For each artist a, quarter t, and language l,

we construct the release share

$$share_{at}^{l} \equiv \frac{\#\text{ releases by artist } a \text{ in quarter } t \text{ in language } l}{\#\text{ total releases by artist } a \text{ in quarter } t}.$$

This panel is used in specifications with quarter fixed effects and artist fixed effects to isolate within-artist changes in language choice over time.

B.3 Construction of Internally Displaced Persons Measures

To account for population displacement and assess robustness to migration-related compositional change, we construct measures of internally displaced persons (IDPs) at the local level using administrative data from the Ukrainian Ministry of Social Policy.

For the period 2014–2019, we use oblast-level counts of registered IDPs provided by the Ministry of Social Policy (upon request of the United Nations). In April 2020, the Ministry publicly reported IDP counts at the district (rayon) level. To obtain a consistent local measure over time, we map post-2020 districts to their corresponding pre-reform units and adjust earlier oblast-level series to the post-2020 district structure using population shares.

A known feature of the IDP registry is that it includes some individuals—particularly pensioners and disabled persons—who formally registered in government-controlled areas to access social payments while continuing to reside in non-government-controlled territories. To mitigate the influence of such registrations on measured displacement intensity, we adjust IDP counts by netting out the estimated share of pensioners and disabled individuals among registered IDPs. Specifically, we compute this share using administrative statistics from seven western regions of Ukraine, where incentives for cross-line registration were minimal due to distance from occupied territories, and apply the resulting adjustment factors to other regions. This procedure follows the logic that registration-based overcounting was most pronounced in areas geographically proximate to the conflict zone.

The resulting IDP measures capture variation in local displacement intensity while reducing sensitivity to administrative registration behavior. In the empirical analysis,

these measures are used as controls and in robustness checks; results are not sensitive to alternative constructions or to excluding IDP adjustments altogether.

B.4 Administrative Reform and Geographic Units

Ukraine's 2020 administrative reform consolidated the pre-existing rayon system into a smaller number of larger districts (rayons), eliminating the earlier category of cities of regional significance as separate administrative units. Prior to the reform, Ukraine was divided into 490 rayons and 187 cities of regional significance; after 2020, these units were reorganized into 136 post-reform rayons.

In analyses of the 2014 conflict, we use fixed effects and clustering at the pre-2020 district level (old rayons and cities of regional significance), which corresponds to the administrative units governing respondents' local environment at that time. In analyses of the 2022 invasion, we instead use fixed effects and clustering at the post-2020 district level, consistent with the contemporaneous administrative structure.

A key constraint is that district-level ethnic composition—specifically, the share of ethnic Russians—is observed only in the 2001 Census and is defined for pre-2020 districts. We therefore use this measure throughout the paper. For post-2020 observations, we map respondents' settlements to their corresponding pre-2020 districts, allowing us to merge the census-based ethnic composition measure and include it (and its interactions with post-2022 indicators) alongside post-2020 district fixed effects.

C Additional Results

Table C.1: The Effect of the 2014 War on Ethnic Identity: Five-Category Gradient

	(1) Only Ukrainian	(2) More Ukrainian	(3) Mixed Ukr./Rus.	(4) More Russian	(5) Only Russian
Panel A: Post-war shift					
Post-February 2014	3.753**	-0.133	-1.798**	-0.256	-1.566***
	(1.527)	(0.693)	(0.903)	(0.457)	(0.468)
Dependent variable mean	79.782	6.491	6.862	2.898	3.967
Dependent variable st. dev.	40.163	24.637	25.282	16.775	19.518
Observations	87,272	87,272	87,272	87,272	87,272
Controls	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Panel B: Heterogeneity by pre-war share of ethnic Russians					
Post-February 2014 × share of ethnic Russians	55.669***	1.000	-27.472***	-8.576**	-20.622***
	(11.594)	(4.479)	(9.378)	(3.514)	(3.718)
Dependent variable mean	79.782	6.491	6.862	2.898	3.967
Dependent variable st. dev.	40.163	24.637	25.282	16.775	19.518
Observations	87,272	87,272	87,272	87,272	87,272
Controls	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓	✓

Notes: This table reports regression estimates of respondents' reported ethnic identity on a post–February 2014 indicator and its interaction with pre-war local ethnic composition, testing the polarization hypothesis. The dependent variable is a five-category identity gradient ranging from exclusively Ukrainian to exclusively Russian (with intermediate categories capturing more Ukrainian, mixed Ukrainian–Russian, and more Russian). Panel A reports coefficients on *Post–February 2014*, an indicator equal to one for interviews conducted after February 2014. Panel B reports coefficients on *Post–February 2014 × Share of ethnic Russians*, where the share is measured at the district level (pre-2020 rayons and cities of regional significance) using the 2001 Census. All specifications include district fixed effects; Panel B specifications additionally include survey-wave fixed effects. All columns include controls for sex, age, education, an urban/rural indicator, district-level inflows of internally displaced persons per capita, regional GDP per capita, and CPI. Standard errors (in parentheses) are clustered at the district level.

Table C.2: The Effect of the 2014 War on Language Use: Three-Category Gradient

	(1) Replying in Ukrainian	(2) Replying in Mixed (Ukr.–Rus.)	(3) Replying in Russian
Panel A: Post-war shift			
Post-February 2014	1.835	-1.079	-0.756
	(2.332)	(1.115)	(2.042)
Dependent variable mean	47.479	13.282	39.239
Dependent variable st. dev.	49.937	33.938	48.829
Observations	95,117	95,117	95,117
Controls	✓	✓	✓
Rayon FE	✓	✓	✓
Panel B: Heterogeneity by pre-war share of ethnic Russians			
Post-February 2014 × share of ethnic Russians	-10.628**	24.542***	-13.914**
	(4.653)	(6.212)	(5.808)
Dependent variable mean	47.479	13.282	39.239
Dependent variable st. dev.	49.937	33.938	48.829
Observations	95,117	95,117	95,117
Controls	✓	✓	✓
Rayon FE	✓	✓	✓
Wave FE	✓	✓	✓

Notes: This table reports regression estimates of respondents' interview-language choice on a post–February 2014 indicator and its interaction with pre-war local ethnic composition, testing the polarization hypothesis. The dependent variables are indicators for replying in Ukrainian, replying in mixed Ukrainian–Russian, and replying in Russian (columns 1–3). Panel A reports coefficients on *Post–February 2014*, an indicator equal to one for interviews conducted after February 2014. Panel B reports coefficients on *Post–February 2014 × Share of ethnic Russians*, where the share is measured at the district level (pre-2020 rayons and cities of regional significance) using the 2001 Census. All specifications include district fixed effects; Panel B specifications additionally include survey-wave fixed effects. All columns include controls for sex, age, education, an urban/rural indicator, district-level inflows of internally displaced persons per capita, regional GDP per capita, and CPI. Standard errors (in parentheses) are clustered at the district level.

Table C.3: The Effect of the 2022 Invasion on Language Use: Three-Category Gradient

	(1) Replying in Ukrainian	(2) Replying in Mixed (Ukr.–Rus.)	(3) Replying in Russian
Panel A: Post-war shift			
Post-February 2022	19.667***	0.041	-19.708***
	(1.440)	(0.640)	(1.837)
Dependent variable mean	59.068	10.584	30.348
Dependent variable st. dev.	49.171	30.763	45.977
Observations	48,744	48,744	48,744
Controls	✓	✓	✓
Rayon FE	✓	✓	✓
Panel B: Heterogeneity by pre-war share of ethnic Russians			
Post-February 2022 × share of ethnic Russians	53.366***	35.040***	-88.406***
	(9.172)	(2.445)	(9.321)
Dependent variable mean	59.068	10.584	30.348
Dependent variable st. dev.	49.171	30.763	45.977
Observations	48,744	48,744	48,744
Controls	✓	✓	✓
Rayon FE	✓	✓	✓
Wave FE	✓	✓	✓

Notes: This table reports regression estimates of respondents' interview-language choice on a post–February 2022 indicator and its interaction with pre-war local ethnic composition, testing the polarization hypothesis. The dependent variables are indicators for replying in Ukrainian, replying in mixed Ukrainian–Russian, and replying in Russian (columns 1–3). Panel A reports coefficients on *Post–February 2022*, an indicator equal to one for interviews conducted after February 2022. Panel B reports coefficients on *Post–February 2022 × Share of ethnic Russians*, where the share is measured at the pre-2020 district level (pre-2020 rayons and cities of regional significance) using the 2001 Census. All specifications include post-2020 district fixed effects (post-2020 rayons). For internally displaced respondents, the district is defined as the pre–February 2022 district of residence; Panel B specifications additionally include survey-wave fixed effects. Specifications control for sex, age, education, an urban/rural indicator, pre-2020 district share of ethnic Russians, and an indicator for being internally displaced interacted with the survey-wave identifier. Standard errors (in parentheses) are clustered at the post-2020 district level.

Table C.4: Spotify Popularity by Region (19/07/2020 - 19/02/2022)

Region	SpotifyPopularity
Kyiv city	100
Kharkiv Oblast	97
Lviv Oblast	90
Odessa Oblast	73
Dnipropetrovsk Oblast	66
Cherkas'ka oblast	66
Khmel'nyts'ka oblast	65
Zaporiz'ka oblast	65
Chernivets'ka oblast	64
Ivano-Frankivs'ka oblast	64
Rivnens'ka oblast	63
Khersons'ka oblast	62
Zakarpats'ka oblast	62
Mykolaivs'ka oblast	62
Vinnyts'ka oblast	60
Poltavs'ka oblast	60
Kirovohrads'ka oblast	59
Chernihivs'ka oblast	58
Ternopil's'ka oblast	58
Kyivs'ka oblast	58
Volyns'ka oblast	58
Sevastopol' city	57
Zhytomyrs'ka oblast	55
Sums'ka oblast	54
Luhans'ka oblast	52
Donetsk Oblast	52
Crimea	51

Notes: The table reports regional Google Trends search interest in Spotify for the period July 19, 2020 to February 19, 2022. Values are normalized on a 0–100 scale within Ukraine, where 100 corresponds to the region with the highest relative search intensity over the selected period.

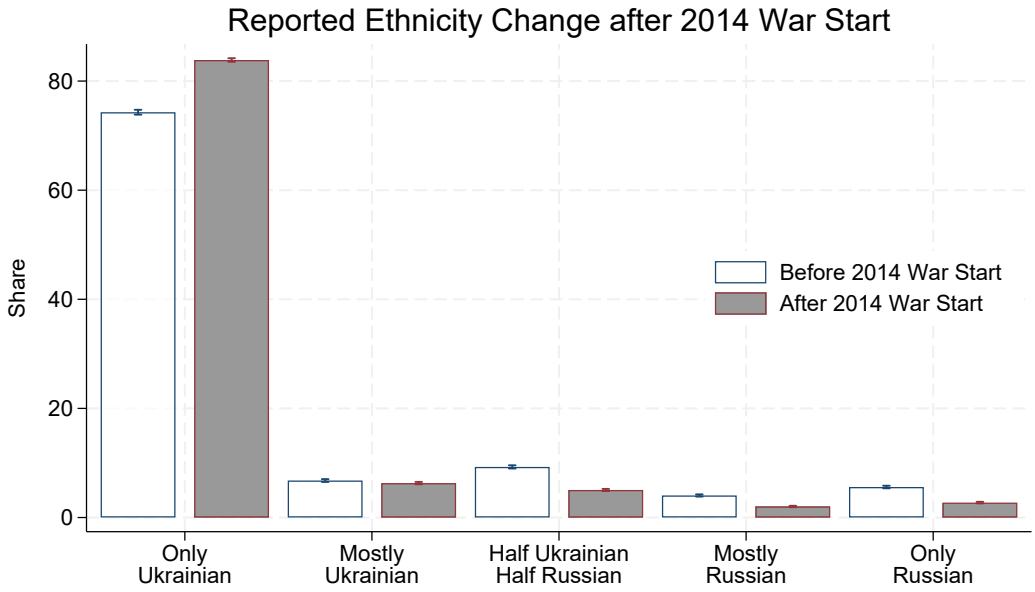


Figure C.1: Reported ethnic identity before and after the 2014 conflict onset

Notes: This figure plots mean shares of respondents across five self-reported ethnicity categories in the KIIS survey, comparing the pre-conflict period (2008–2013) to the post-conflict period (2014–2020). The five categories are: only Ukrainian, mostly Ukrainian, mixed Ukrainian–Russian, mostly Russian, and only Russian. Bars show period means computed after excluding observations from Crimea and districts that were occupied after 2014 (fully or partially). Error bars show 95% confidence intervals.

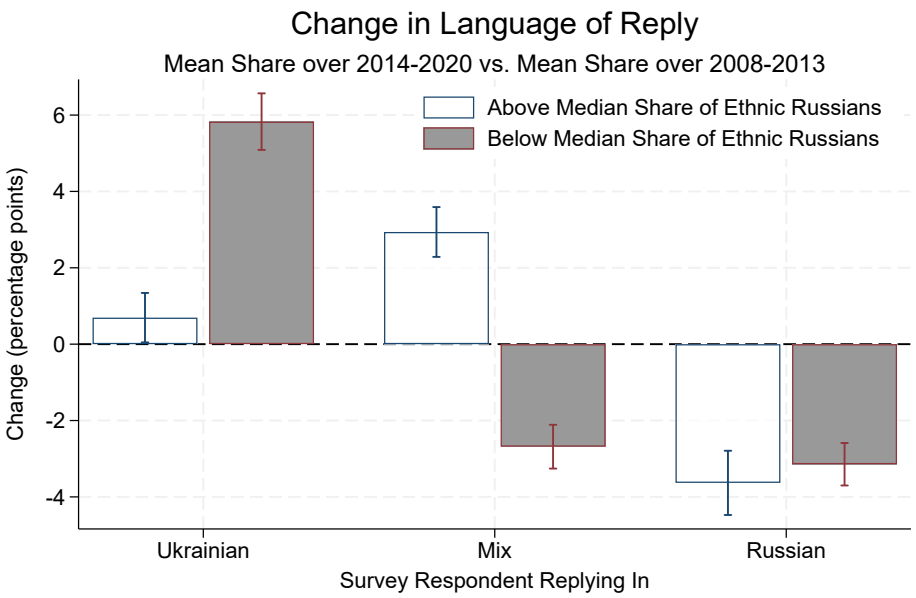
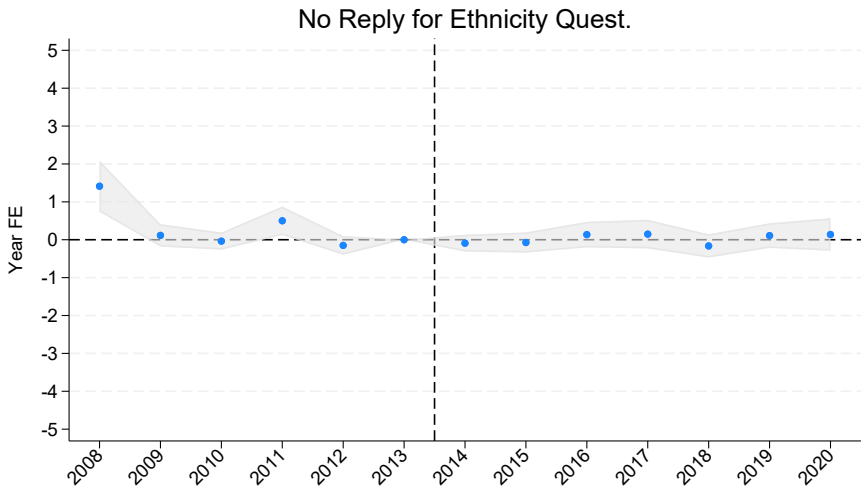


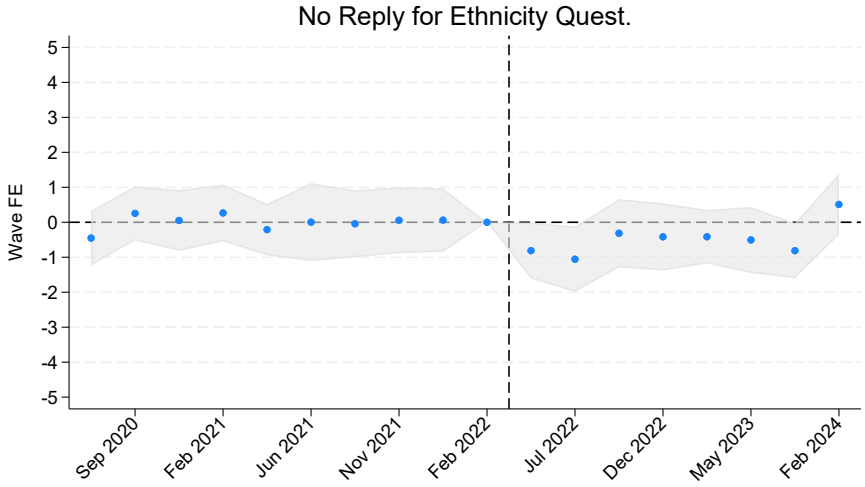
Figure C.2: Change in Language of Reply Post 2014 War, by Pre-War Ethnic Composition

Notes: This figure plots changes in the mean share of KIIS respondents replying in Ukrainian, mixed Ukrainian–Russian, and Russian between the post-war period (March 2014–2020) and the pre-war period (2008–February 2014). The sample excludes observations from Crimea and districts that were fully or partially occupied after 2014. Bars report $\Delta\bar{x} = \bar{x}_{2014–2020} - \bar{x}_{2008–2013}$, expressed in percentage points. White bars correspond to respondents in districts with an above-median pre-war share of ethnic Russians measured using the 2001 Census; gray bars correspond to respondents in districts with a below-median share. Error bars show 95% confidence intervals for the difference in period means.

Figure C.3: No-reply ethnicity question: time fixed effects



(a) Year fixed effects around the 2014 conflict



(b) Survey-wave fixed effects around the 2022 invasion

Notes: This figure plots the estimated time fixed effects from regressions of an indicator for not answering the ethnicity question in surveys. Panel A reports yearly fixed effects (normalized to 2013) for the 2014 conflict period, controlling for pre-2020 district (rayon/city of regional significance) fixed effects and individual covariates (sex, age, education, and an urban/rural indicator). Panel B reports survey-wave fixed effects (monthly; normalized to February 2022) for the 2022 full-scale invasion period, controlling for post-2020 district (rayon) fixed effects and individual covariates (age, sex, education, cohort-by-five-year fixed effects, urban/rural, family finances). Standard errors are clustered at the district level (pre-2020 districts in Panel A; post-2020 districts in Panel B). The figure is used to assess social-desirability bias by showing whether item nonresponse on ethnicity shifts discontinuously around the two conflict episodes.



Figure C.4: Attitudes Toward Russian-Speaking Ukrainians, 2020–2023

Notes: The figure reports the distribution of responses to the question of whether respondents would accept Russian-speaking Ukrainians in various social roles over 4 waves between 2020 and 2023. Social roles include: as family members, close friends, neighbors, work colleagues, residents of Ukraine, or guests of Ukraine, as well as the share stating they would not let them into Ukraine. Each bar sums to 100 percent within survey wave.

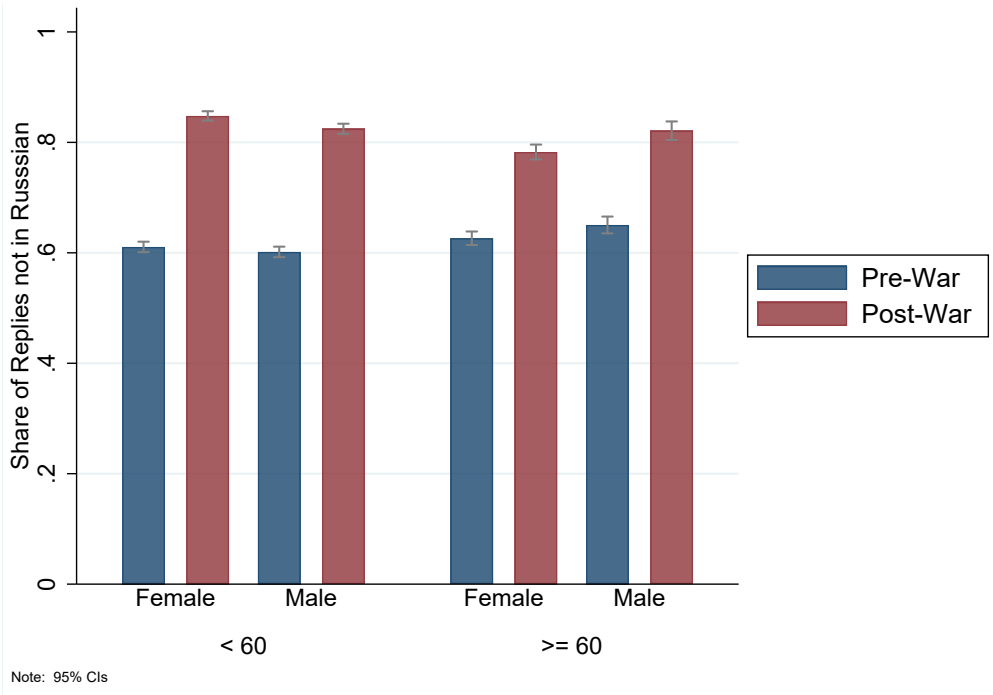


Figure C.5: Change in Ukrainian Language Use in Survey Interviews by Sex and Age

Notes: The figure shows the share of survey respondents replying not in Russian (equal to one for replying in Ukrainian or mixed Ukrainian–Russian) before and after the February 2022 full-scale invasion. Estimates are reported separately by sex and age group (below 60 and 60 or older). Bars display group means and vertical lines indicate 95% confidence intervals.

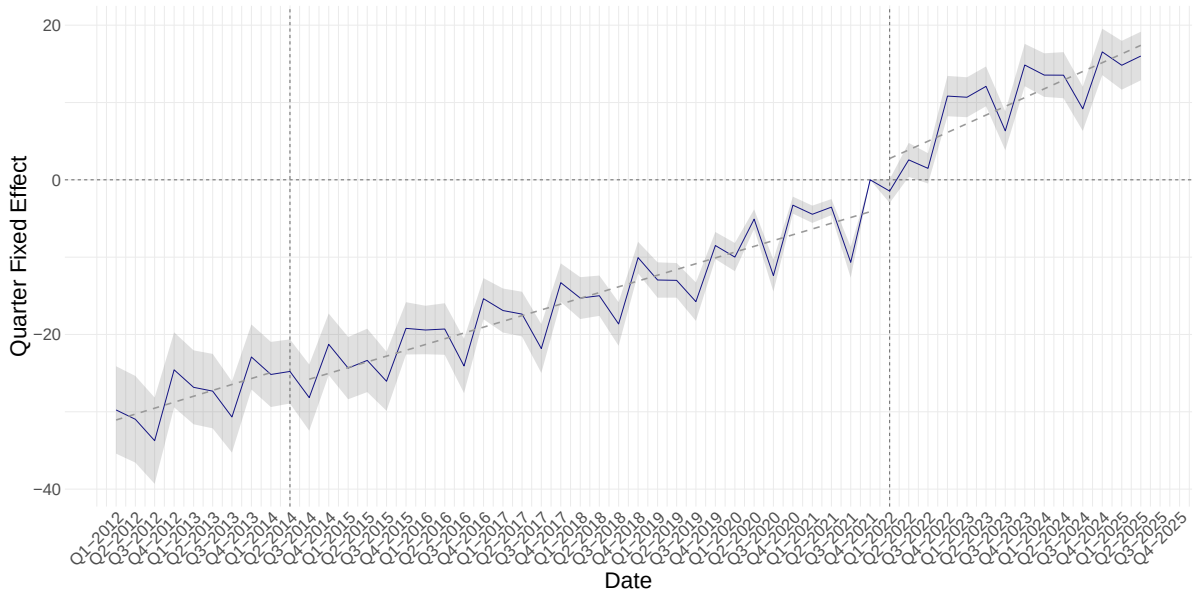


Figure C.6: Quarterly Fixed Effects for the Share of Google Searches in Ukrainian

Notes: The figure plots quarterly time fixed effects, where the outcome is the share of Ukrainian spelling in total google searches for word i in region r and quarter t. The sample includes 198 Ukrainian–Russian word pairs and excludes occupied and partially occupied regions. Observations are weighted by word–region search volume as described in the text. Standard errors are clustered at the word–region level; whiskers denote 95% confidence intervals.

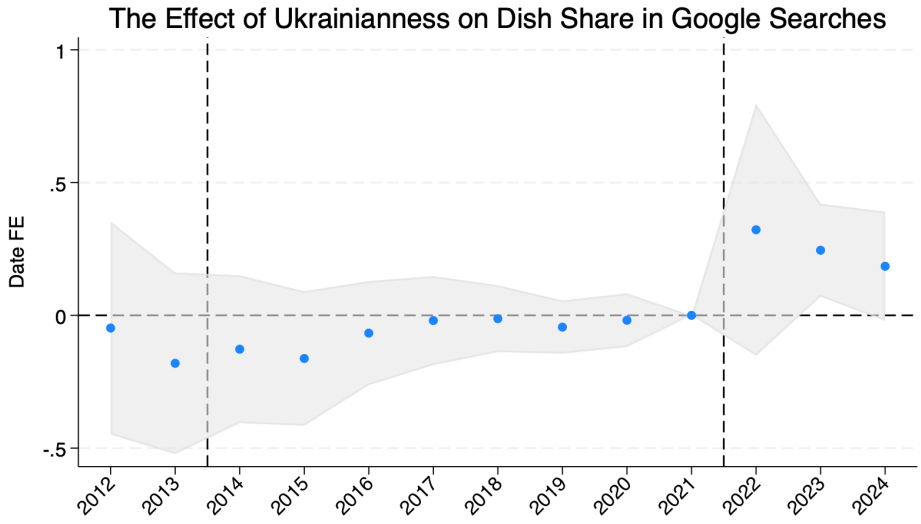


Figure C.7: Cuisine searches: effect of dish Ukrainianness on dish share (standardized outcome).

Notes: The figure reports yearly event-study coefficients from the specification 4 but using the standardized outcome $Pop^{std}_{irt} = \frac{Pop^{raw}_{irt} - \overline{Pop^{raw}_{i,2012\text{–}2021}}}{sd(Pop^{raw}_{i,2012\text{–}2021})}$. Coefficients therefore capture changes in the Ukrainianness gradient relative to 2021 after removing dish-specific scale differences. The model includes dish-by-oblast and oblast-by-month fixed effects. Standard errors are two-way clustered by dish and oblast; whiskers denote 95% confidence intervals.

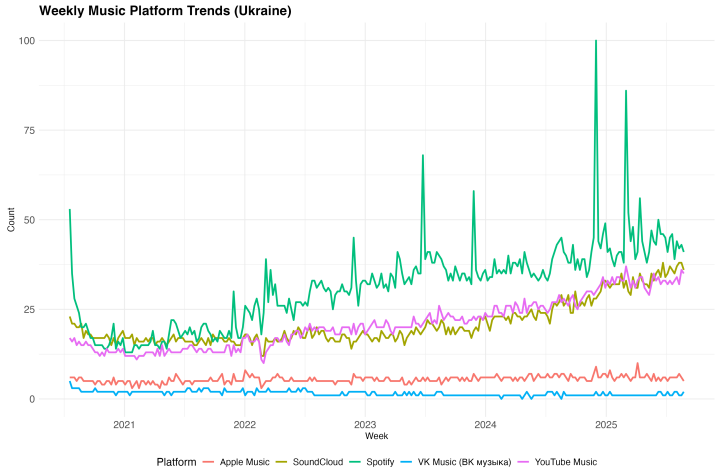


Figure C.8: Google Search Interest in Music Streaming Platforms in Ukraine, 2020–2025

Notes: The figure plots Google Trends search interest in Ukraine for major music streaming platforms between July 19, 2020 and August 30, 2025. Values are Google Trends indices scaled to 0–100 within each query, where 100 denotes peak search interest over the selected period.

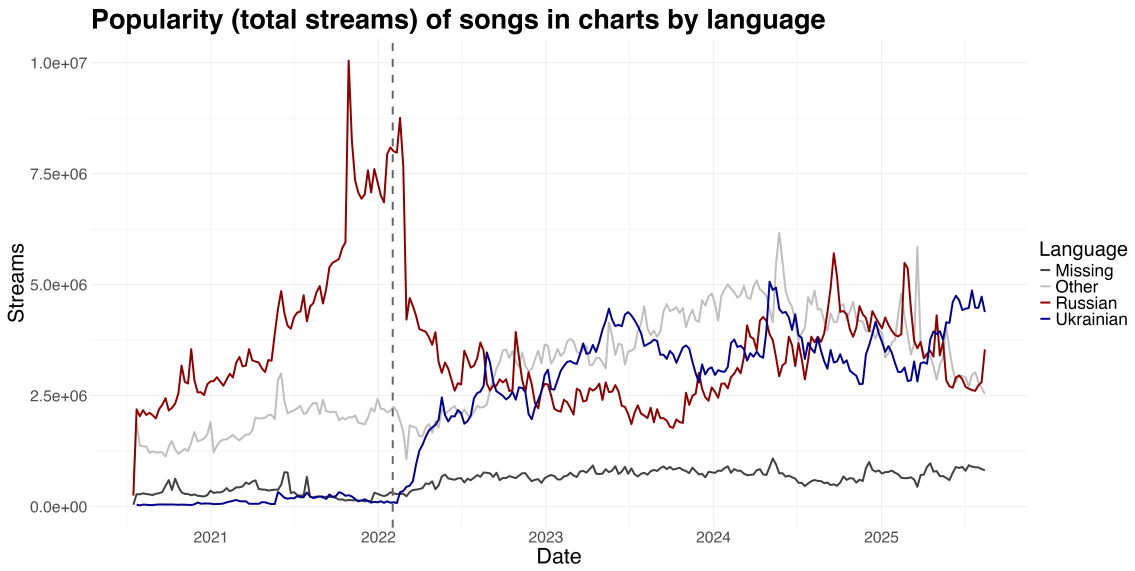
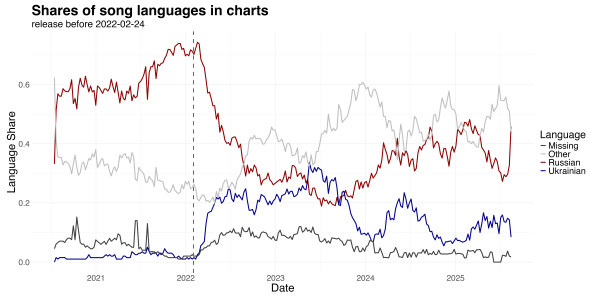
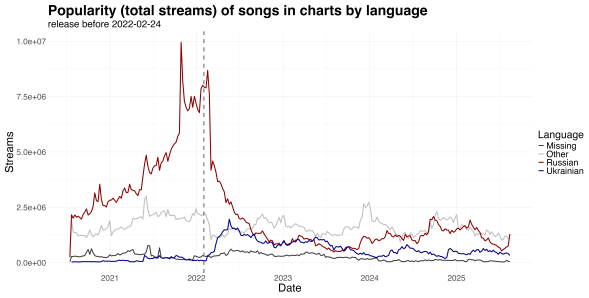


Figure C.9: Spotify Weekly Top Songs Ukraine: Total Weekly Streams by Track Language, 2020–2025

Notes: The figure reports, at the weekly frequency, the total number of streams in the Spotify Weekly Top Songs Ukraine chart by language (Ukrainian, Russian, or other). Track language is determined using lyric-based language detection from Genius; tracks without available lyrics (e.g., instrumentals or songs not documented on Genius) are classified separately when relevant.



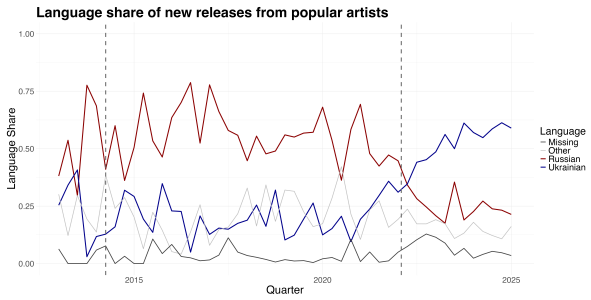
(a) Share of chart entries by language (pre-2022 releases).



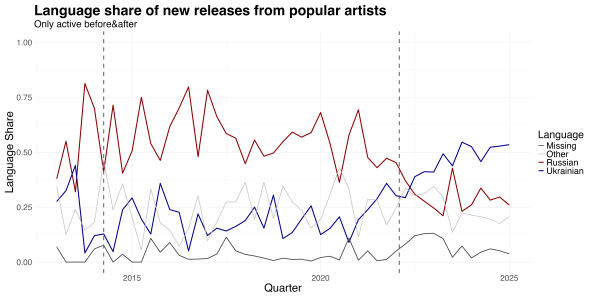
(b) Total weekly streams by language (pre-2022 releases).

Figure C.10: Spotify Weekly Top Songs Ukraine: Language Composition for Tracks Released Before February 24, 2022

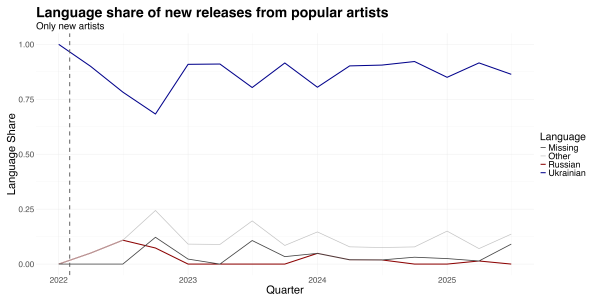
Notes: The figure replicates the language composition analysis restricting the sample to tracks released prior to February 24, 2022. Panel (a) shows the weekly share of chart entries by language; Panel (b) aggregates total weekly streams by language group. Track language is assigned using lyric-based detection from Genius. Because the set of songs is fixed to pre-invasion releases, post-2022 changes reflect demand-side reallocation rather than the introduction of new music.



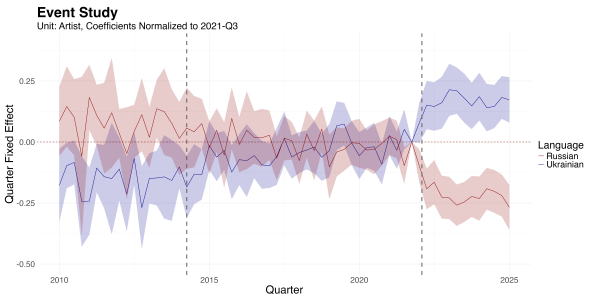
(a) All releases (Ukrainian artists).



(b) Releases by artists active before and after Feb. 24, 2022.



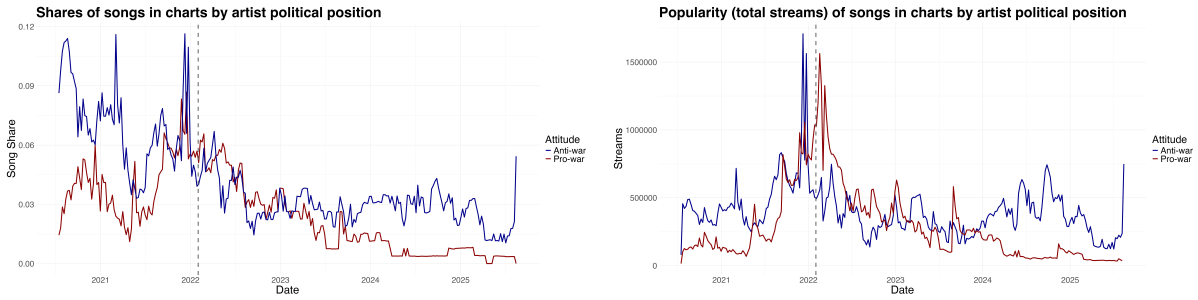
(c) Releases by new artists (first release after Feb. 24, 2022).



(d) Within-artist language shift (artist-level event study).

Figure C.11: Supply-Side Response of Ukrainian Artists: Language of Releases Before and After February 24, 2022

Notes: The figure documents changes in the language of music releases by artists of Ukrainian origin. Panels (a)–(c) plot quarterly shares of releases by language (Ukrainian, Russian, other) for: (a) all Ukrainian artists in the sample; (b) artists who released both before and after Feb. 24, 2022 ("active" artists); and (c) artists whose first release occurs after Feb. 24, 2022 ("new" artists). Panel (d) isolates within-artist changes by estimating an artist fixed-effects regression at the artist–quarter level, where the outcome is the share of an artist's releases in a given language and the coefficients are quarter-by-year time effects normalized to the pre-invasion period; standard errors are clustered at the artist level and caps denote 95% confidence intervals. Across panels, the post-2022 increase in Ukrainian-language releases reflects both the entry of predominantly Ukrainian-language artists and language switching among continuing artists.

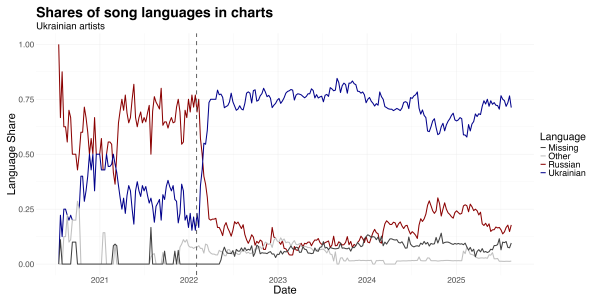


(a) Share of chart entries by artist stance.

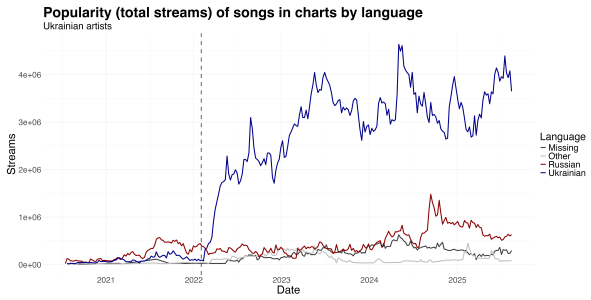
(b) Total weekly streams by artist stance.

Figure C.12: Spotify Weekly Top Songs Ukraine: Popularity of Russian Artists by Stance on the War, 2020–2025

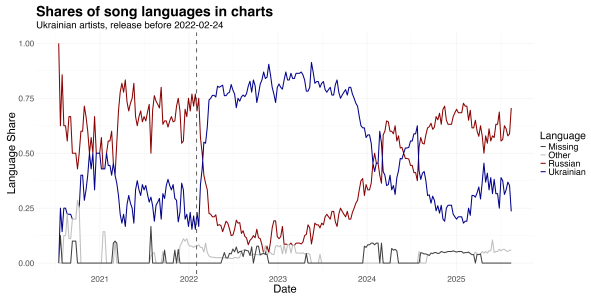
Notes: The figure tracks the popularity in Ukraine of songs by Russian artists classified by their public stance on the war using the Stars About War Project (2025). Panel (a) shows the weekly share of chart entries accounted for by pro-war and anti-war artists; Panel (b) aggregates total weekly streams for songs by each group. The underlying data are the Spotify Weekly Top Songs Ukraine charts (July 17, 2020–August 14, 2025). Artist stance is available for a subset of charting Russian artists (18 pro-war and 21 anti-war among artists appearing at least once).



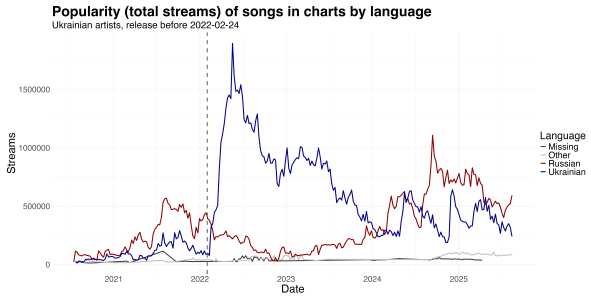
(a) Share of chart entries (all releases).



(b) Total weekly streams (all releases).



(c) Share of chart entries (pre-2022 releases).



(d) Total weekly streams (pre-2022 releases).

Figure C.13: Popularity of Songs by Ukrainian Artists Before and After the Full-Scale Invasion

Notes: The figure reports, at the weekly frequency, the share of chart entries and total streams in the Spotify Weekly Top Songs Ukraine chart for songs performed by artists of Ukrainian origin. Panels (a) and (b) use all charting songs. Panels (c) and (d) restrict the sample to songs released before February 24, 2022. Shares are computed over chart entries; streams are aggregated within week by language (Ukrainian, Russian, other). Because Panels (c) and (d) fix the set of songs to pre-invasion releases, post-2022 changes in those panels reflect shifts in relative demand rather than the introduction of new music.