War and Nation-Building: Evidence from Ukraine*

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

October 2026

Abstract

How quickly and how deeply can wars reshape national identity? We study this question in the context of Ukraine and the Russia–Ukraine war, using rich data spanning ethnic, linguistic, and cultural dimensions of identity. We document large and persistent shifts toward Ukrainian identity following the onset of the war in 2014 and Russia's full-scale invasion in 2022. Nationally representative quarterly surveys comprising over 150,000 observations show substantial increases in Ukrainian ethnic self-identification after both episodes. Language adjustment was modest after 2014, but after 2022 the share replying in Russian fell by about 20 percentage points, corroborated by within-user language changes on Twitter and by shifts toward Ukrainian in Google searches. Consistent with national consolidation rather than polarization, the largest shifts occurred in districts with historically higher shares of ethnic Russians. Cultural consumption and intergenerational choices have also shifted sharply since 2022, toward Ukrainian cuisine, Ukrainian-language music, and more Ukraine-associated baby names. A range of evidence argues against compositional change or population displacement as primary explanations. Together, the results show that 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 prewar 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, Marcella Alsan, Ruben Enikolopov, Vicky Fouka, Ruixue Jia, Danielle Lussier, Ben Olken, Panu Poutvaara, Nancy Qian, Tavneet Suri, Marco Tabellini, Nico Voigtländer, Romain Wacziarg, Jaya Wen, Ekaterina Zhuravskaya, and conference participants at the AEA Annual Meeting, the ifo Workshop on Reconstruction after Conflict, the LSE–Kiel–CEPR Conference on Defence Economics, the MPSA Politics and Political Economy of Eurasia Workshop, NBER Economics of Culture and Institutions, and the SIOE Annual Conference for helpful comments. We thank Olga Podbutskaya, Andrey Samarskiy, Timofey Shvaya, and especially Aleksander Vedishchev for their superb research assistance. We are grateful to Valerio Sergio Castaldo for his contributions at earlier stages of the project.

†Rome Economic Doctorate (e-mail: sabramenko@redphd.it).
‡UPF, Barcelona School of Economics, 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 King's College London (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

When Russia occupied and 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. We show that, a decade later, the opposite pattern has emerged. Rather than fragmenting, Ukrainian national identity has consolidated, with the largest shifts occurring precisely in districts where ethnic Russians were most concentrated before the war. This consolidation intensified following Russia's full-scale invasion in 2022, generating large and persistent shifts in subjective ethnic identification, language use, and cultural consumption.

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, 2025). Whether external war forges nations or fractures them is a classical question in political thought (Tilly, 1975)—and the answer is theoretically ambiguous. In models of endogenous social identity (Shayo, 2009; Sambanis and Shayo, 2013), identification depends on the relative status of competing groups, the salience of different group boundaries, and individuals' perceived distance from group prototypes. External conflict can raise the salience and status of the nation, strengthening national identification (Sambanis et al., 2015). But it can also activate subnational cleavages—particularly when internal minorities are culturally close to the adversary. For example, in the case of the 2000–2004 Israeli-Palestinian conflict, Palestinian attacks led to persistent polarization between Israeli Arabs and Israeli Jews within Israel (Shayo and Zussman, 2011, 2017). We develop a framework that organizes these forces and yields testable predictions about the conditions under which war produces nation-building rather than internal division.

Ukraine provides an especially sharp setting for studying the role of war in nation-building because, prior to 2014, polarization was unusually plausible. A substantial share

of the population was ethnolinguistically close to the aggressor state: in 2012, roughly 30% reported some Russian ethnic identity, and a majority used Russian or a Ukrainian–Russian mix in daily communication. Political cleavages were also pronounced, with public opinion over European integration sharply divided and with pro-Russian parties remaining electorally significant as late as 2021. Russian cultural and information products were widely consumed, further blurring national boundaries. Together, these features made Ukraine close to a knife-edge case: external aggression could have consolidated the nation, but it could also have activated preexisting ethnolinguistic cleavages and produced internal fragmentation. 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 the two major war shocks—annexation of Crimea and start of the Donbas war in 2014 and Russia's full-scale invasion in 2022—reshaped national identity in Ukraine. Our framework predicts stronger consolidation after 2022, especially in costly behaviors such as language use, as both the salience of the Ukraine–Russia divide and Ukraine's relative standing increased more sharply. We examine this by combining 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 their onset.[1] Second, we implement a difference-in-differences design that leverages continuous variation in prewar ethnolinguistic composition across districts, measured by the share of ethnic Russians in the 2001 Census. This second approach directly distinguishes consolidation from polarization by testing whether identity shifts following the onset of the conflict are larger or smaller in areas with stronger historical ties to Russia.

We begin by examining changes in self-reported ethnic identification. Following the onset of hostilities in 2014, the share of respondents identifying as *only* Ukrainian rises by approximately 7 percentage points, with corresponding declines in Russian and mixed

[1]We rely on substantial evidence—both historical and based on high-stakes economic outcomes—that the 2014 war episode was unexpected (Treisman, 2018; Silva and Volkova, 2018; Korovkin and Makarin, 2023; Korovkin et al., 2026). Although some Western governments anticipated a 2022 invasion, the scale of the shock was also largely unexpected, as reflected in stock-market reactions and the absence of stockpiling in Russia's imports of dual-use products (Federle et al., 2022; Egorov et al., 2026). As late as February 19, 2022, President Zelensky expressed skepticism about an imminent invasion (BBC News Russian, 2022).

identification. After Russia's full-scale invasion in 2022, the share reporting Ukrainian ethnicity increases by approximately 6 percentage points, reaching 96%, while only about 2.5% continue to report Russian or mixed ethnicity.

Difference-in-differences estimates point to national consolidation rather than polarization: after both shocks, the shift toward Ukrainian ethnic self-identification is systematically larger in districts with higher prewar shares of ethnic Russians. Moving from a district with no ethnic Russians to a district with 25% of ethnic Russians is associated with an additional increase in Ukrainian identification of about 16 percentage points after the onset of the 2014 war and 10 percentage points after the 2022 invasion. Event-study estimates show sharp breaks at both conflict onsets and no evidence of differential pretrends across districts, lending support to the identifying assumptions underlying our research design.

We next turn to language use, a behavioral outcome whose adjustment requires effort and carries social and cognitive costs. Here the contrast between the two conflict episodes is especially stark. Changes after 2014 are modest. By contrast, the 2022 full-scale invasion produces a large and abrupt shift toward Ukrainian. In survey interviews, the share of respondents answering in Ukrainian or a Ukrainian–Russian mix rises by approximately 20 percentage points following the invasion—a strikingly large change along such a costly margin. As with ethnic identification, difference-in-differences estimates show that the shift is largest in districts with higher prewar concentrations of ethnic Russians.

The contrast between the two war phases is consistent with our framework. Both episodes accentuated the Ukraine–Russia divide, especially in 2022, favoring consolidation. The status channel also favored consolidation more strongly in 2022, when Ukraine's effective resistance, Russia's military setbacks, and international support plausibly raised Ukraine's relative standing. The framework therefore predicts stronger consolidation after 2022, especially on costly behavioral margins such as language use.

Evidence from additional data sources corroborates this pattern and argues against compositional change and social desirability as primary explanations. Using original tweets geolocated to Ukraine, we document a 15-percentage-point increase in Ukrainian-language use *within* individual accounts. Comparing Ukrainian and Russian versions of the same high-frequency words in Google searches, we find an 8- to 10-percentage-point increase in the Ukrainian-language share, indicating that the change extends to private

settings in which social-desirability concerns are unlikely to play a central role.

Cultural outcomes spanning music, cuisine, and baby-naming choices provide further evidence that identity change extends across diverse behaviors. Using Spotify's weekly charts of the top-200 songs in Ukraine, we document a large and remarkably rapid reorientation of music consumption toward Ukrainian-language content following the 2022 invasion. The share of Ukrainian-language songs in the chart rises from about 2% before the invasion to nearly 30% afterward and remains elevated through 2025. Because music consumption on Spotify is largely private, this pattern is also unlikely to reflect interviewer demand or social-desirability concerns. We further show that the shift reflects changes in both listener demand and artist behavior, is not driven by boycotts of Russian artists, persists even when restricting attention to Ukrainian artists, and is unlikely to be explained by listener compositional change or population displacement.

Beyond music, we observe a parallel reorientation in both everyday consumption and durable household choices. Google searches shift by roughly 0.2–0.3 standard deviations toward Ukrainian rather than Russian cuisine, while published lists of popular baby names shift toward names more strongly associated with Ukrainian than Russian naming traditions. A similar shift appears in the names given to ethnic Ukrainian children in Kazakhstan, whose parents face few of Ukraine's wartime incentives to adopt Ukrainian-coded names, suggesting that the pattern is not driven solely by strategic adaptation to Ukraine's domestic environment. Taken together, these patterns show that conflict-induced identity change extends beyond stated preferences and public-facing behavior to private consumption choices and durable decisions with intergenerational consequences.

We then probe the channels highlighted by our conceptual framework. First, the shift toward Ukrainian identity reflects more than rejection of the Russian out-group: rising Ukrainian identification is accompanied by declines not only in Russian identification but also in other, non-Russian ethnic identification. Within our framework, this broad-based positive pull is consistent with an increase in the relative status of the Ukrainian nation. Second, several findings point to the salience channel. Districts experiencing more war-related violence between survey waves exhibit sharper declines in Russian-language replies. Further, although attitudes toward Russians living in Ukraine do deteriorate after 2022, the decline is far more pronounced for attitudes toward Russians living in

Russia. This divergence suggests that the invasion heightened the salience of the interstate boundary between Ukraine and Russia more than Ukraine's internal ethnolinguistic divide. Third, the limited available data suggest that where Russia swiftly established control and Ukrainian resistance was limited, such as in Crimea and parts of Donbas occupied in 2014, identification shifted less toward Ukraine and may even have shifted toward Russia—underscoring that consolidation is contingent on the course of the conflict.

Finally, we show that Ukrainian identity is closely related to willingness to sustain resistance. Relative to otherwise comparable Russian-language respondents, those interviewed in Ukrainian are 14.6 percentage points less likely to support territorial concessions in exchange for peace, against a sample mean of about 20%. The association is similar for men of draftable age. Although descriptive, the results suggest that national consolidation may matter for a society's resilience and willingness to continue resisting external aggression.

Overall, the results point to a broad-based and persistent shift toward Ukrainian national identity. They show that identity is far more malleable than is often assumed: war can generate large, rapid, and durable changes in identity, even in heterogeneous societies with longstanding ethnolinguistic and cultural ties to the adversary.

Our first contribution is to the literature on conflict and nation-building, which has long asked whether external war forges a common national identity or activates preexisting cleavages and fragments the polity.[2] A body of work shows that war and occupation can durably reshape political and group identities in either direction (Dell and Querubin, 2018; Dehdari and Gehring, 2022; Ananyev and Poyker, 2023). We add by showing that external war can catalyze a large and rapid reorientation of ethnolinguistic and cultural identity in the country under attack. We also show that war can consolidate national identity even where a substantial share of the population has ethnolinguistic ties to the aggressor. Our difference-in-differences design reveals that shifts toward Ukrainian identity are largest where prewar ties to Russia were strongest: consolidation extends across the very cleavage that the invasion could have deepened. Suggestive evidence from occupied territories points to weaker or reversed shifts where Russia rapidly established control, highlighting that consolidation depends on the course of the conflict. Our conceptual framework

[2]The idea that external conflict may strengthen internal cohesion has a long tradition in the social sciences (Simmel, 1904; Coser, 1956); a related literature links war to state formation and national integration (Tilly, 1975). See Appendix A for a review.

rationalizes these contrasting responses and clarifies why external threat may consolidate identity in some settings but deepen internal divisions in others.

Our second contribution is to broaden the empirical measurement of identity change. A growing body of work studies how policies, institutions, historical shocks, and shared experiences shape national identification and language use (Jha, 2013; Aghion et al., 2019; Bazzi et al., 2019; Fouka, 2019, 2020; Depetris-Chauvin et al., 2020; Alesina et al., 2021; Atkin et al., 2021; Carlitz et al., 2025; Weigand et al., 2025; Bazzi et al., 2026; see Rohner and Zhuravskaya, 2025 for a recent overview). Our contemporary setting offers unusually rich data for studying nation-building: we combine high-frequency nationally representative surveys with digital traces and administrative records to track identity change across outcomes that differ in visibility, persistence, and adjustment costs. Survey responses are observed by interviewers, and tweets and baby names are visible to others; conversely, searches and music listening are largely private. The joint movement of these variables therefore helps distinguish internalized identity change from public conformity. Methodologically, our approach illustrates how integrating surveys with large-scale digital traces and administrative records can strengthen inference about identity formation.[3]

The rest of this paper is organized as follows. Section 1 develops a conceptual framework grounding our predictions in a model of endogenous social identity. Section 2 describes the Ukrainian context. Section 3 presents the data sources. Section 4 reports the empirical results for ethnic and linguistic identity. Section 5 examines cultural identity outcomes. Section 6 examines mechanisms. Section 7 suggests possible implications of identity change for national defense. Section 8 concludes.

1 Conceptual Framework

We ground our empirical analysis in a simple model of endogenous social identity developed in Shayo (2009) and applied to conflict settings by Sambanis and Shayo (2013)

[3]Related work on Ukraine examines the war's economic consequences (Korovkin and Makarin, 2023; Korovkin et al., 2026) and refugee flows (Adema et al., 2024). Closest to our analysis, Kulyk (2016, 2018, 2024) uses a few nationwide survey snapshots to document shifts in Ukrainian identity and language use after 2014 and 2022. These studies document the broad pattern of Ukrainization, but their reliance on a limited number of survey waves makes it difficult to isolate discrete breaks at the onset of each conflict from longer-run secular change. We advance this literature by using 72 quarterly survey waves and newly accessed settlement identifiers to trace identity around both conflict shocks, directly test consolidation against polarization across historically more and less Russian districts, and show that the reorientation extends beyond survey-based identity to within-user language choice, private behavior, and durable household decisions.

and Sambanis et al. (2015).[4] Formal derivations are in Appendix B.

1.1 Social Identity and Identity Choice

Consider a resident of Ukraine who belongs to multiple social groups but chooses how much to identify with each. The relevant competition in our setting is between two group identities: the *Ukrainian nation* (N) and a *Russian or Russian-leaning identity* (R). Identification with group $j \in \{N, R\}$ has two components. First, the individual cares about the group's *status*—i.e., its perceived power, success, or prestige—denoted by S_j. Second, the individual seeks to minimize their perceived *distance* from the group's prototype, denoted by d_{ij}. The utility that individual i derives from identifying with group j is

$$I_{ij} = \gamma\, S_j - (1 - \gamma)\, d_{ij} \tag{1}$$

where $\gamma \in (0, 1)$ is the relative importance of group status vs. perceived distance. Thus, to the extent that individual i actually identifies with group j, they seek to promote group j's goals and to conform to group j's norms.

Perceived distance depends on how similar the individual is to the group's prototype along cognitively salient dimensions:

$$d_{ij} = \sum_k w_k \delta_k(i, j) \tag{2}$$

where $\delta_k(i, j)$ denotes the distance between individual i and group j's prototype along attribute k (e.g., ethnicity, language, or cultural practices), and $w_k \geq 0$ is the *attention weight* assigned to that dimension. Attention weights are endogenous: events such as military attacks can make particular attributes more salient, thereby increasing their weight in the individual's perceived distance from each group's prototype.

Let $\alpha_{ij} \geq 0$ denote the weight individual i places on group j, where $\sum_j \alpha_{ij} = 1$. Total utility when individual i places different weights on different identities is then given by

$$U_i = \pi_i + \sum_j g(\alpha_{ij})\, I_{ij} \tag{3}$$

where π_i is material payoff (net of the costs of behavioral adjustment introduced below), $g(0) = 0$, $g' > 0$, and $g'' \leq 0$. From Equation (1), the higher α_{ij} is, the more i cares about

[4] Hauk (2025) extends this framework to foreign interventions in divided societies. For broader reviews, see Shayo (2020); Genicot and Ray (2024); Gidron and Shayo (2026).

group j's status and about their perceived distance from the group's prototype. However, we allow for diminishing marginal utility from identifying with any specific group.[5] We focus on individuals for whom both identities are relevant, and who may thus place a positive weight on either a Ukrainian (α_{iN}) or a Russian-leaning identity (α_{iR}), or both.

Identity is endogenous: the identity weights α_{ij} are choice variables. An individual places more weight on Ukrainian national identity when I_{iN} is high relative to I_{iR}. Three forces therefore govern identification choices: the *relative status* of the groups (S_N vs. S_R), the *salience* of different dimensions (w_k), and the individual's initial attributes, such as language and cultural practices (which determine δ_k).

Beyond choosing how much to identify with each group, the individual can change their behavior, thereby repositioning themselves in attribute space. For each dimension k, let b_{ik} denote a behavioral adjustment toward the national prototype (switching to Ukrainian, consuming Ukrainian cultural goods, choosing Ukrainian names). Movement in the opposite direction is captured by $b_{ik} < 0$. Adjustment is costly: we assume the cost $c_{ik}(b_{ik})$ is convex, and its marginal cost varies across dimensions and across individuals (changing everyday language requires more effort than changing stated ethnic identification; language is easier to change when young). Identifying more strongly with the nation increases the return to moving toward its prototype, hence adjustment is larger for those who identify more with the nation and along dimensions that are less costly to change.[6]

1.2 How External War Shapes Identity

We now apply this framework to the Russia–Ukraine war. The key insight is that the war's effect on national identity is not automatic: it depends on how the conflict shifts status, salience, and distances in attribute space, which in turn depend on the war's nature and outcomes. Note that while in the more general framework, fighting intensity, group

[5] This nests the simple linear, discrete-choice case $g(\alpha) = \alpha$, in which individuals generically identify with a single group (the optimum is a corner) and comparative statics concern population shares. Appendix B.4 derives the comparative statics under $g'' < 0$.

[6] Formally, the individual chooses the identity weights and the behavioral adjustments to solve

$$\max_{\alpha_{iN} \in [0,1],\, \{b_{ik}\}} \; U_i = \pi_i^0 + g(\alpha_{iN})\, I_{iN} + g(1-\alpha_{iN})\, I_{iR} - \sum_k c_{ik}(b_{ik})$$

where the distance entering I_{ij} in (1) and (2) is the effective gap after adjustment, $\delta_k(i,N) = \delta_k^0(i,N) - b_{ik}$ and $\delta_k(i,R) = \delta_k^0(i,R) + b_{ik}$, with $\delta_k^0(i,j)$ being the baseline (preadjustment) gap. π_i^0 is material payoff before adjustment costs, so that in (3) $\pi_i = \pi_i^0 - \sum_k c_{ik}(b_{ik})$. Appendix B analyzes this problem.

status, and the salience of the national vs. ethnic divide may themselves be endogenous to (aggregate) individual and political choices, in this paper we seek to understand individual responses to exogenous shifts in status and salience following the two invasion episodes.

Status. Military resistance and international solidarity plausibly raise the perceived status of the Ukrainian nation (S_N). Conversely, Russian success increases the status associated with a Russian identity (S_R) whereas Russian military setbacks and international isolation lower it. A higher relative national status increases the appeal of Ukrainian identification for all residents, but most consequentially among individuals for whom both identities are relevant, especially ethnic Russians and Russian speakers.

The status effects likely differ across the two conflict episodes. In 2014, Russia achieved major territorial objectives at relatively low cost: it annexed Crimea with little fighting, while Russian-backed forces established control over parts of Donbas. Ukraine mobilized, but was unable to reestablish control over these territories. Russia's rapid gains and the limited international response likely raised its relative status. The 2022 full-scale invasion produced the opposite dynamics. Russia's military setbacks, Ukraine's effective resistance, and strong international support raised Ukrainian status while lowering Russia's.[7]

Salience. War shifts the attention weights placed on different group boundaries. The critical distinction is between *interstate* salience (Ukraine vs. Russia) and *intrastate* ethnolinguistic salience (divisions within Ukraine by ascribed ethnicity and by language). By attacking the country as a whole, including cities where Russian was widely spoken, the invasions made the interstate boundary the dominant divide. In terms of the framework, this increased the weight placed on citizenship and country of residence relative to the markers that divide Ukrainians internally, above all ascribed ethnicity (see Section 2). Holding attribute distances (δ_k's) fixed, this reduces d_{iN} especially for Russian-speaking Ukrainians, whose attention shifts away from the attributes that separated them from the national prototype, while increasing d_{iR} by placing greater weight on attributes that distinguish them from the Russian prototype.[8] The direction of this effect is not a property

[7]Survey evidence supports this contrast: the share of Ukrainians very proud of their citizenship rose by 40 percentage points after the 2022 invasion, but by only 8 percentage points after 2014 (DIF survey, 2025). These shifts likely also reflect salience shocks, to which we now turn.

[8]The framework does not require the weight on language itself to fall. What matters is the shift of attention toward the interstate boundary. The weight on language may rise, fall, or stay the same (see

of conflict-induced changes in salience as such. Reversing the reallocation of attention reverses the identity response: a conflict that made the *within*-Ukraine ethnolinguistic boundary dominant—under discriminatory targeting or different patterns of exposure—could move identification the other way (see Shayo and Zussman, 2011 for evidence). Which response occurs depends on how the war is fought and on where residents stand on the attributes it makes salient.

While the salience shock had the same sign across the two episodes, its scale was much larger in 2022. In 2014, hostilities were geographically confined to Crimea and parts of Donbas; in 2022, the full-scale invasion brought attacks to cities throughout Ukraine, including Kyiv, Kharkiv, and Odesa, affecting the whole country.

1.3 Predictions

This framework yields three sets of predictions that guide our empirical analysis.

Prediction 1: Consolidation. Under the conditions of the 2022 invasion, where the aggressor attacks the country as a whole and Ukrainian resistance raises national status, the framework predicts a broad shift toward Ukrainian identification. More generally, any identity shift, in either episode and direction, is concentrated among individuals for whom both identities are relevant. Districts with higher prewar shares of ethnic Russians contain more such individuals, so the observed shift should be larger there.[9] Consolidation therefore implies larger shifts toward Ukrainian identity in higher-Russian-share districts, whereas polarization implies smaller shifts or movement in the opposite direction.

Prediction 2: Behavioral change. The shift toward Ukrainian identification should manifest not only in stated ethnicity but also in identity-linked behaviors that reduce distance from the national prototype: language use, cultural consumption, and naming choices.[10]

Appendix B.4).

[9] Beyond the composition effect (more responsive individuals), the district gradient may also reflect a larger per-person shift. The size of the individual response depends on the marginal identification return g', which is larger for individuals less anchored in the Ukrainian national identity, while individuals at a corner do not respond at all (see Appendix B.4). Higher-Russian-share districts contain more individuals who are both responsive and less firmly anchored.

[10] This is unambiguous for identity shifts due to status shocks. It is also unambiguous for identity shifts due to salience shocks for behaviors whose own salience is unchanged. However, where the war also reduced the attention weight placed on a given behavior, the lower weight would work against adjustments along that behavioral margin. Depending on whether the individual identifies more strongly with the nation or the Russian-leaning identity, this can reinforce or counterbalance the identity effect. The opposite obtains for a behavior whose salience increased. See Appendix B.4.

Prediction 3: Differential effects across episodes. Following the discussion above, the status and salience channels reinforce one another in 2022 but work in opposite directions in 2014. Thus, the direction of the 2014 identity response is an empirical question. At the same time, the framework predicts a stronger and unambiguous shift toward Ukrainian national identity after 2022. The contrast between the two episodes should be more pronounced on costly behavioral margins such as language, either because starting to change is itself costly (i.e., the cost function has a kink at zero) so a small identity shift does not trigger it, or because small changes are not detected in our empirical measures (see Appendix B).

1.4 Persistence and Contingency

The analysis above is static. It describes how identification and behavior respond jointly to given levels of status and salience. Whether the changes we document persist after the war is an open question; answering it requires modeling how status, salience, and adjustment costs evolve, which is beyond the scope of this paper. But the framework does suggest mechanisms that could generate persistence. First, some adjustments are durable or costly to reverse. A child's name tends to be fixed, and learning or habitually using Ukrainian lowers the future cost of using it. Second, the cultural environment may itself change. The expansion of Ukrainian-language music (which we document in Section 5.1) may lower the cost of consuming Ukrainian culture, and the national prototype may shift. Third, stronger national identification may increase willingness to resist and mobilize (we explore this link in Section 7). In equilibrium, this could raise national status and thereby sustain consolidation. Such feedback can support multiple equilibria (Sambanis and Shayo, 2013), so that a temporary shock could move a society from one equilibrium to another.

More broadly, our framework implies that war's effects on identity depend on how the conflict reshapes status, salience, and differences in attributes. In Ukraine, the combination of a broad-based external attack, Russia's military setbacks, effective Ukrainian resistance, and international support likely shifted these forces toward national consolidation. Yet, under different military and political conditions, the same forces could produce polarization rather than nation-building. In Section 6.3, we provide suggestive evidence for this contingency using identity changes in territories occupied by Russia in 2014.

2 Ukrainian Context

2.1 Ethnic and Linguistic Identity in Ukraine

In Ukraine, the Soviet concept of *nationality* is commonly used in place of ethnicity, as reflected in the 2001 census and the surveys we analyze. Under Soviet rule, nationality was determined by parental origin and recorded as a mandatory attribute in civil passports. We thus refer to self-reported *nationality* as ethnic self-identification. High rates of ethnic intermarriage in Ukraine (25% in 1989; Rapawy, 1997), particularly in the east and south, contribute to the fluidity of ethnic self-identification, making Ukraine a fertile setting for identity change (Jia and Persson, 2021).

Language is another important dimension of identity in Ukraine. From the mid-nineteenth century onward, Russification shaped education, media, and public life (Kravtsiv and Kubijovyc, 1993). The Valuev Circular of 1863 banned most Ukrainian-language publications, and the Ems Ukaz of 1876 extended these restrictions to education and theater (Subtelny, 2009, 282–284). Yet Ukrainian remained the predominant spoken language in rural areas and a marker of local identity (Magocsi, 1996, 479–480). In the 1920s, Soviet Ukrainization briefly reversed this trajectory: by 1929, more than 80% of schools taught in Ukrainian. Stalin ended this interlude in 1933 (Subtelny, 2009, 389–390, 422). Russification deepened over the following decades. The Holodomor and World War II sharply reduced Ukraine's population, and large-scale Russian resettlement altered its ethnic and linguistic composition (Subtelny, 2009, 415–416, 480–485; Markevich et al., 2025). By the Brezhnev era, Russian became central to higher education and economic mobility. The share of books published in Ukrainian fell from 60% in 1958 to 24% in 1980 (Subtelny, 2009, 536).[11] Many ethnic Ukrainians adopted Russian as their primary language. Even after independence, Russian remained prevalent in media, education, and daily communication. As of 2001, 87.8% of the population were fluent in Ukrainian and 65.7% in Russian (Ukrstat, 2001).

From 2014 onward, Moscow repeatedly invoked the need to "protect" ethnic Russians and Russian speakers to justify its aggression (Le Monde, 2024). Russian political narratives denied the existence of a distinct Ukrainian nation and sought to activate subnational

[11] Another feature of this linguistic landscape is *Surzhyk*, a mixed Ukrainian–Russian vernacular long stigmatized as "incorrect" speech. Since the 2022 invasion, it has experienced a revival as an inclusive marker of Ukrainian identity (Osypenko, 2025).

identities among Russian-speaking Ukrainians, particularly in the east and south, through appeals to shared origins and claims of linguistic repression (Bækken, 2024).

Following Arel and Khmelko (1996), we find it useful to distinguish three broad ethnolinguistic groups: Ukrainian-speaking Ukrainians, Russian-speaking Ukrainians, and predominantly Russian-speaking ethnic Russians. The latter two groups are especially informative for distinguishing consolidation from polarization: conflict could either draw them toward a common Ukrainian identity or reinforce Russian-oriented identification. Importantly, ethnolinguistic composition varied sharply across space before the war: Russian ethnicity and language were more prevalent closer to Russia and along the Black Sea coast, with substantial variation across districts even within the same region (Appendix Figure D.1). This spatial variation allows us to examine whether identity change was greater in areas with larger prewar shares of ethnic Russians and whether prewar geographic differences subsequently narrowed, a feature we exploit throughout the analysis.

2.2 A Brief Account of the Russia–Ukraine War (2014–2026)

The Outbreak (2014–2015). Mass protests began in November 2013 after President Yanukovych declined to sign the EU–Ukraine Association Agreement. In February 2014, the protests culminated in a change of government, Yanukovych fled to Russia, and Ukraine's parliament (the Verkhovna Rada) assumed executive authority. The following month, Russia annexed Crimea after deploying armed forces and holding an internationally unrecognized referendum. Soon afterward, Russian-backed groups seized administrative buildings in Donetsk and Luhansk and proclaimed "people's republics." Although Moscow denied direct involvement, Russian military units participated in key engagements, including at Ilovaisk and Debaltseve. Ukraine mobilized its forces but was unable to restore control over the occupied areas (Plokhy, 2015, ch. 27; D'Anieri, 2023, ch. 7).

Military Stalemate (2015–2022). The February 2015 Minsk II agreement established a ceasefire but left the central political and territorial disputes unresolved. Russian-backed forces retained control over roughly one-third of the Donetsk and Luhansk oblasts (regions), including their regional capitals. Although large-scale fighting subsided, lower-intensity violence persisted, with periodic surges in its intensity. Implementation of the agreement

stalled, and the frontline remained largely unchanged through early 2022 (Plokhy, 2015, ch. 27; D'Anieri, 2023, ch. 8).

The Full-Scale Invasion (2022–2026). On 24 February 2022, Russia launched a full-scale invasion of Ukraine. Military operations initially targeted Kyiv and other major cities, but by April, Russian forces had withdrawn from northern Ukraine following sustained Ukrainian resistance. The war subsequently shifted to the east and south, where Russian forces occupied additional territory in the Donetsk and Luhansk oblasts and advanced into the Zaporizhzhia and Kherson oblasts. In September 2022, Russia claimed all four oblasts in their entirety, despite not exercising full control over them. Ukrainian counteroffensives recaptured most Russian-held territory in Kharkiv oblast in September and the city of Kherson in November, while subsequent operations in 2023 produced more-limited territorial changes. By early 2024, the war had largely become one of attrition, characterized by sustained fighting along relatively stable frontlines (D'Anieri, 2023, ch. 9; CFR, 2024).

Consequences. The war has imposed enormous economic and human costs. The 2014 war reduced economic activity by an estimated 40% to 50% in areas controlled by Russian-backed forces (Kochnev, 2019) and by about 6% to 7% elsewhere in Ukraine through disrupted production networks and trade (Korovkin and Makarin, 2023; Korovkin et al., 2026). Following the 2022 invasion, Ukraine's GDP contracted by about 31% (Federle et al., 2026). By the end of 2025, cumulative military casualties were estimated at 600,000 for Ukraine and over 1.2 million for Russia (Jones and McCabe, 2026). Around 3.6 million Ukrainians were internally displaced and another 6.8 million had sought refuge abroad, while military strikes caused extensive damage to energy and transport infrastructure (CFR, 2024; The UN Refugee Agency, 2025). Systematic quantitative evidence on the war's cultural and identity consequences, however, is scant; our paper seeks to fill this gap.

3 Data

To quantify the identity responses to the war episodes, 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 self-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.

3.1 Survey Data: KIIS Omnibus Surveys

Our primary data source on ethnic identification and language choice is the omnibus survey conducted by the Kyiv International Institute of Sociology (KIIS). We use 72 survey waves fielded roughly 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, we obtained access to restricted KIIS records identifying respondents' city or settlement of residence, allowing us to link individuals to prewar local ethnolinguistic composition at a fine geographic level. This feature enables within-country heterogeneity analysis that is infeasible using the coarser regional identifiers available in publicly released KIIS data.

We focus on three survey-based identity outcomes. First, respondents answer an open-ended question: "What is your nationality, as you consider yourself?" Second, respondents who identify as ethnically Ukrainian and/or Russian place themselves on an ordered five-point scale capturing mixed ethnic identities.[12] This second measure is available through early 2022. Third, interviewers record the language of the interview (Ukrainian, Russian, or mixed) following a standardized protocol in which they offer respondents a choice of language through a bilingual prompt. This measure reflects revealed linguistic choice in a low-stakes interaction. Further details on the data and question wording are provided in Appendix C.1.

3.2 Twitter Data: Within-Individual Changes in 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. Although geotagging became less common over time, 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

[12]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.

identification comes from within-user changes in language choice over time. This design holds user composition fixed within the estimation sample and provides behavioral validation of the survey evidence in a public communication setting.

3.3 Google Search Data: Private Language Use and Consumption

We complement survey and social media evidence with Google Trends data, which capture search behavior. Because search activity is not interviewer-mediated and is predominantly private, these data are less susceptible to social-desirability or strategic reporting concerns. In 2021, Google held 92% of Ukraine's search market (StatCounter, 2021).

We use Google Trends to measure two dimensions of identity-related behavior. First, we construct quarterly measures of the relative use of Ukrainian and Russian in queries from 2012 to 2025 for each oblast (Ukraine's first-level administrative region). These measures, which are based on pairs of Ukrainian and Russian spellings of the same words, capture revealed language choice outside the survey setting (see Appendix C.2.1 for details).

Second, we use search activity to study cultural interest, focusing on national cuisine. We assemble an oblast-by-month panel measuring search intensity for major Ukrainian and Russian dishes from 2012 to 2024. We construct a "Ukrainianness" score for each dish based on its relative search prevalence in Ukraine vs. Russia before 2022. We then examine whether search activity shifts differentially toward dishes more strongly associated with Ukraine following the 2022 invasion (see Appendix C.2.2 for details).

3.4 Music Consumption: Spotify Charts

To directly analyze cultural consumption, we study the market for popular music. We combine Spotify's weekly charts of the top-200 songs in Ukraine from July 2020 to August 2025, scraped from the music-data aggregator Kworb, with song-level metadata from the Spotify Web API, and lyrics retrieved from Genius and classified by language.[13] Furthermore, for analyzing supply-side responses, we merge these data with a comprehensive catalog of Ukrainian recording artists and their full discographies; see Appendix C.4 for details. The resulting dataset contains 3,850 distinct charting tracks by 572 artists, with information on language, release date, chart trajectory, and total streams. This structure allows us to distinguish changes in listener demand from artists' language choices.

[13] Genius is an online platform hosting song lyrics and crowdsourced annotations.

3.5 Baby Names: Intergenerational Identity

To explore effects on more-durable intergenerational identity choices, we use administrative data from the Ukrainian Ministry of Justice listing the most popular baby names by oblast from 2015 onward. Each name is assigned a Ukrainianness score based on its relative prewar prevalence in Ukraine vs. Russia, using data from *forebears.io*. We then assess whether relatively more-Ukrainian names become more popular after the 2022 invasion, controlling for oblast-by-time fixed effects. As an external validation, we use administrative birth records from the Bureau of National Statistics of Kazakhstan (2026) to examine naming choices among ethnic Ukrainian children born in Kazakhstan over 2019–2025. See Appendix C.3 for details.

3.6 Prewar Ethnolinguistic Composition

To measure preexisting proximity to Russia, we use ethnolinguistic composition from Ukraine's latest census, conducted in 2001. Our main measure is the share of residents reporting Russian ethnicity in each pre-2020 rayon (Ukraine's second-level administrative district). Using the share of Russian native speakers yields similar results.

Ukraine's 2020 administrative reform consolidated rayons into larger units. To preserve a consistent measure of treatment intensity, we map post-2020 survey respondents back to their corresponding prereform districts based on their settlement of residence (for waves after February 2022, the settlement in which they lived before the invasion) and use the 2001 composition measure throughout. This approach allows us to combine prewar ethnolinguistic composition with contemporaneous district fixed effects and clustered standard errors. See Appendix C.5 for additional details.

3.7 Population Displacement

To assuage concerns about compositional changes 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, adjusting for administrative changes using the rayon-level data for 2020. For the period following the 2022 full-scale invasion, we rely on rayon-level IDP dynamics from the International Organization for Migration (IOM). See Appendix C.6 for details.

4 Results: Ethnic and Linguistic Identity

In this section, we examine whether the 2014 war and the 2022 full-scale invasion consolidated Ukrainian national identity or instead polarized the population along ethnolinguistic lines. We first present survey-based evidence on ethnic self-identification and interview language, which provide direct measures of identity and its behavioral expression. Then, we turn to Twitter and Google search data to assess whether the same patterns extend beyond the survey setting to within-user changes in public language use, and to language choice in private settings.

4.1 Baseline Outcomes: Ethnicity and Language in the KIIS Survey

We use the 2008–2024 KIIS data to trace how respondents' ethnic self-identification and interview language changed around the two major phases of the war. This analysis provides an initial test of Predictions 1 and 3 from Section 1.

4.1.1 Empirical Strategy

We proceed in two steps. First, we document the time dynamics of ethnic self-identification and interview language around each war phase. Second, we estimate difference-in-differences specifications that exploit cross-district variation in prewar ethnic composition. Specifically, we ask whether districts with larger Russian minorities experience weaker or stronger shifts toward Ukrainian identity following conflict onset.

To study average postwar changes in identity, we estimate for respondent i in rayon r and survey-wave (quarter) t:

$$\text{Identity}_{irt} = \beta \, \text{PostWar}_t + X'_{irt}\delta + \mu_r + \varepsilon_{irt} \tag{4}$$

We then examine heterogeneity by local ethnic composition using difference-in-differences:

$$\text{Identity}_{irt} = \beta \, (\text{PostWar}_t \times \text{EthnicRussians}_{r,2001}) + X'_{irt}\delta + \mu_r + \gamma_t + \varepsilon_{irt} \tag{5}$$

In both equations, Identity_{irt} denotes one of three outcomes: (i) an indicator for identifying ethnically as *only Ukrainian*, rather than mixed or Russian, in the 2014 analysis; (ii) an indicator for identifying ethnically as Ukrainian in the 2022 analysis; or (iii) an indicator for responding in Ukrainian or a Ukrainian–Russian mix rather than in Russian. PostWar_t is an indicator for survey waves after February 2014 in the 2014 analysis, and after February

2022 in the 2022 analysis. $\text{EthnicRussians}_{r,2001}$ is the share of ethnic Russians in rayon r in the 2001 Census. X_{irt} contains individual- and locality-level controls (detailed below), while μ_r and γ_t denote rayon and survey-wave fixed effects. For the 2022 analysis, fixed effects and clustering use post-2020 rayons, while the ethnic-Russian share is measured at the pre-2020 rayon level, with its main effect included in X_{irt} in every specification. Respondents are assigned to these geographic units based on their preinvasion residence. For the 2014 analysis, fixed effects and clustering use pre-2020 rayons.

In our conceptual framework, these outcomes are discrete proxies for latent variables that we do not directly observe: the identity weight α_{iN} and the behavioral adjustments b_{ik}. What we observe are ethnic self-identification for α_{iN} and behavioral indicators such as interview language for b_{ik}. If z^* is the relevant latent variable, we can think of the observed indicator as $\mathbf{1}\{z^* \geq \tau\}$, where τ is the threshold for switching on the indicator. The linear specifications in (4) and (5) are linear-probability approximations to this mapping.

In Equation (4), β captures the average postconflict change in identity outcomes. Identification requires that, absent the respective war phase, average identity outcomes would not have shifted systematically between the pre- and postwar periods, conditional on controls and rayon fixed effects. Both episodes were large, abrupt shocks, limiting the scope for anticipatory adjustment: the onset of hostilities in 2014 was unexpected (Treisman, 2018; Silva and Volkova, 2018; Korovkin and Makarin, 2023; Korovkin et al., 2026), and although some Western governments anticipated a possible invasion in 2022, its scale appears to have been largely unanticipated (Federle et al., 2022; Egorov et al., 2026).[14]

In Equation (5), β captures how postwar identity shifts vary with the prewar share of ethnic Russians. Positive estimates ($\beta > 0$) indicate stronger shifts toward Ukrainian identity in areas with larger prewar shares of ethnic Russians, which we interpret as evidence of *consolidation*. Negative estimates ($\beta < 0$) indicate *polarization*. Identification relies on parallel trends: absent the conflict, rayons with different prewar shares of ethnic Russians

[14]The culmination of Euromaidan and the change of government immediately preceded Russia's 2014 aggression, so their effects cannot be fully separated. Several features nevertheless favor an interpretation centered on the conflict. First, the 2022 invasion, which was not accompanied by a change of government, was followed by a similar shift toward Ukrainian identification and a substantially sharper language response. Second, the 2004 Orange Revolution was accompanied by little aggregate change in ethnic identification or home language use (Pop-Eleches and Robertson, 2018). Third, our event studies show no corresponding differential shifts during the Euromaidan protest period beginning in November 2013.

would have experienced similar changes in identity outcomes. We assess the plausibility of this assumption by examining preconflict trends in event-study specifications.

Controls. We include covariates to account for sample composition changes. Our baseline specifications control for sex, age, educational attainment, and urban residence. For the 2014 analysis, we also control for the rayon-level ratio of IDP inflows to population. For the 2022 analysis, we instead observe and control directly for respondents' displacement status.[15]

4.1.2 The 2014 War

Ethnicity. Panel A of Table 1 estimates overall post-2014 changes in identification patterns using Equation (4). Columns (1) and (2) show a large increase in the share of individuals identifying as exclusively ethnically Ukrainian: 8.9 percentage points with rayon fixed effects and 7.1 percentage points after adding individual-level and IDP controls.

Panel B of Table 1 shows the difference-in-differences specification from Equation (5). Columns (1) and (2) provide strong evidence of identity consolidation: the increase in exclusive Ukrainian ethnic self-identification is substantially larger in rayons with higher prewar shares of ethnic Russians. The interaction term *Post* $\times$ *Share of Ethnic Russians* is positive and highly significant. Quantitatively, moving from a rayon with 0% ethnic Russians to a rayon with 25% ethnic Russians increases the likelihood of a shift toward identifying ethnically as *only Ukrainian* after the 2014 war by about 16 percentage points.

Appendix Table D.1 extends both analyses to the full Likert scale of ethnic identity. The results show that the rise in exclusively Ukrainian ethnic self-identification is accompanied by corresponding declines in Russian and mixed identities across the scale.[16]

Panel A of Figure 1 presents event-study estimates from a wave-by-wave version of Equation (5), with exclusively Ukrainian identification as the outcome. The estimates are not significantly different from zero before February 2014, increase as the war unfolds, and remain generally elevated from 2015 onward.

[15] We do not control for local economic conditions as they are themselves affected by the conflict. Nevertheless, adding oblast-by-year consumer price index and GDP per capita to the 2014 specifications changes the difference-in-differences estimates only modestly; corresponding data for the 2022 analysis are unavailable.

[16] For the corresponding raw-data patterns, absent any controls, see Appendix Figures D.2 and D.3.

Table 1: Ethnic and Linguistic Identity Before and After the 2014 War

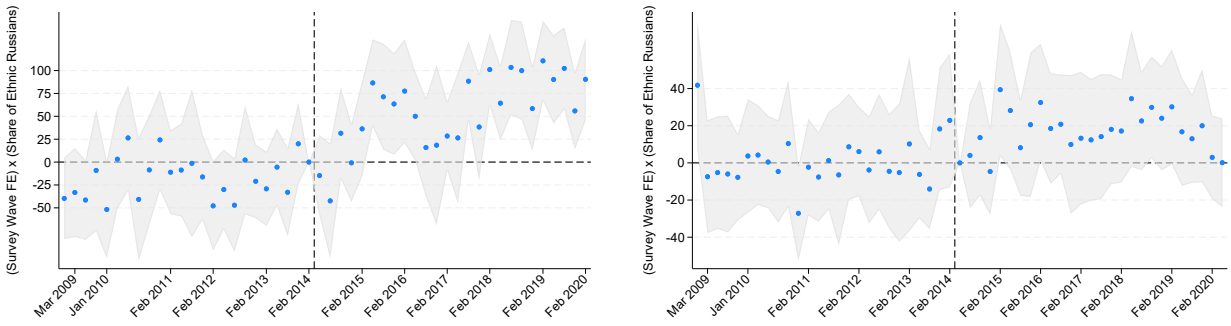
	Identified as Ukrainian Only (%)		Replying Not in Russian (%)	
	(1)	(2)	(3)	(4)
Panel A: Postwar Shift				
Post–February 2014	8.896***	7.127***	2.122**	1.299
	(1.309)	(1.051)	(0.923)	(1.157)
Controls		✓		✓
Rayon FE	✓	✓	✓	✓
Panel B: Difference-in-Differences				
Post–February 2014 × share of ethnic Russians	71.038***	63.761***	16.215***	7.747
	(9.090)	(9.744)	(4.525)	(5.564)
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 examines changes in ethnic and linguistic identity before and after the 2014 war's onset. Dependent variables are indicators multiplied by 100. In columns (1) and (2), the dependent variable equals one if the respondent reports exclusively Ukrainian ethnicity, rather than Russian or mixed. In columns (3) and (4), it equals one if the interview was conducted in a language other than Russian (Ukrainian or a Ukrainian–Russian mix). Panel A reports coefficients on 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 rayon level using the 2001 Census and varies from zero to one. All specifications include rayon fixed effects; Panel B specifications also include survey-wave fixed effects. Columns (2) and (4) also control for sex, age, education, an urban/rural indicator, and rayon-level IDP inflows per population. Dependent-variable summary statistics and numbers of observations are identical across the panels. Standard errors in parentheses are clustered at the rayon level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Language. Columns (3) and (4) of Table 1 report analogous results for the interview language. Once individual- and rayon-level controls are included, there is little statistically significant evidence of either an overall increase in Ukrainian or mixed Ukrainian–Russian replies or a stronger shift in rayons with higher prewar shares of ethnic Russians. Event-study estimates in Panel B of Figure 1 become somewhat positive as the war unfolds, but are imprecise and rarely individually distinguishable from zero.[17]

Overall, the 2014 results reveal a marked shift in ethnic self-identification but only modest adjustments in language use. This pattern is consistent with our conceptual framework: the war sharpened the Ukraine–Russia divide, favoring consolidation through

[17]For the corresponding raw-data patterns, absent any controls, see Appendix Figure D.4.



(a) Ethnicity (Ukrainian only)

(b) Language (replying not in Russian)

Figure 1: The 2014 War: Event-Study by Prewar Ethnic Composition

Notes: This figure plots event-study coefficients from a regression of the indicated identity outcome on survey-wave indicators interacted with the prewar share of ethnic Russians in the respondent's rayon, measured in the 2001 Census (Equation 5). In Panel A the outcome is an indicator for reporting exclusively Ukrainian ethnicity; in Panel B the outcome is an indicator for replying in a language other than Russian (Ukrainian or a Ukrainian–Russian mix). In both panels, positive post-2014 coefficients indicate relatively larger increases in Ukrainian identification and non-Russian replies in rayons with higher prewar shares of ethnic Russians, consistent with national consolidation. 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 rayons that were fully or partially occupied after 2014. Standard errors are clustered at the rayon level. Shaded areas indicate 95% confidence intervals.

the salience channel, while Russia's initial military success likely worked in the opposite direction through the status channel. That the net response was toward consolidation indicates that the salience channel dominated. As a result, self-reported ethnicity, which requires little behavioral adjustment, responds to this more moderate shock, whereas language use—a costlier and more behaviorally demanding margin—adjusts less. The 2022 full-scale invasion provides a setting to examine whether a stronger war shock, where both channels reinforce each other, may lead to broader linguistic realignment.

4.1.3 The 2022 Invasion

Table 2 repeats the structure of Table 1, but self-reported ethnicity (columns 1 and 2) is measured slightly differently, since the postinvasion KIIS waves no longer include the five-category gradient question. We therefore use the survey's direct measure of self-reported ethnicity, which records whether respondents identify as Ukrainian, Russian, mixed Ukrainian–Russian, or other ethnicities.

Panel A documents a large aggregate shift toward Ukrainian ethnic self-identification after February 2022, amounting to 5.6–5.9 percentage points (columns 1 and 2). Panel B shows that this shift is significantly stronger in rayons with higher prewar shares of ethnic

Table 2: Ethnic and Linguistic Identity Before and After the 2022 Invasion

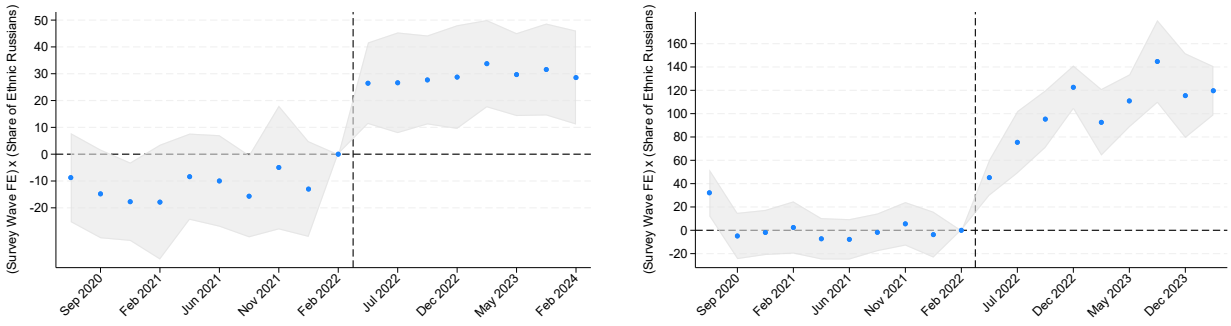
	Identified as Ukrainian (%)		Replying Not in Russian (%)	
	(1)	(2)	(3)	(4)
Panel A: Postwar Shift				
Post–February 2022	5.867***	5.640***	21.180***	19.708***
	(0.682)	(0.675)	(1.809)	(1.837)
Controls		✓		✓
Rayon FE	✓	✓	✓	✓
Panel B: Difference-in-Differences				
Post–February 2022 × share of ethnic Russians	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 examines changes in ethnic and linguistic identity before and after the 2022 invasion. Dependent variables are indicators multiplied by 100; coefficients are in percentage points. In columns (1) and (2), the dependent variable equals one if the respondent reports Ukrainian ethnicity, and zero if they report Russian, both Russian and Ukrainian, or another ethnicity. In columns (3) and (4), it equals one if the interview was conducted in a language other than Russian (Ukrainian or a Ukrainian–Russian mix). Panel A reports coefficients on 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 rayon level using the 2001 Census. All specifications include rayon fixed effects; Panel B specifications also include survey-wave fixed effects. Panel A specifications omit wave fixed effects because they would absorb the *Post–February 2022* indicator. Columns (2) and (4) also control for sex, age, education, an urban/rural indicator, and an indicator for being internally displaced after February 2022. All columns control for the pre-2020 rayon-level share of ethnic Russians, which varies from zero to one. Dependent-variable summary statistics and numbers of observations are identical across panels. Standard errors in parentheses are clustered at the rayon level. See Appendix C.5 for details on the geographic units. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Russians: the interaction coefficient is positive and highly significant with and without controls. 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 stronger shift in the probability of identifying as Ukrainian.

Figure 2 shows the event-study estimates. The estimates in Panel A exhibit no systematic preinvasion trend and rise sharply thereafter, with no evidence of reversion through 2024.[18] Taken together, these patterns point to national consolidation rather than polarization following the 2022 invasion, echoing the pattern observed after the 2014 war.

[18]For the corresponding raw-data patterns, absent any controls, see Appendix Figure D.5.



(a) Ethnicity (Ukrainian) (b) Language (replying not in Russian)

Figure 2: The 2022 Invasion: Event-Study by Prewar Ethnic Composition

Notes: This figure plots event-study coefficients from a regression of the indicated identity outcome on survey-wave indicators interacted with the prewar share of ethnic Russians in the respondent's pre-2020 rayon, measured in the 2001 Census (Equation 5). In Panel A, the dependent variable equals one if the respondent reports Ukrainian ethnicity, and zero if they report Russian, both Russian and Ukrainian, or another ethnicity; in Panel B, it equals one if the interview was conducted in a language other than Russian (Ukrainian or a Ukrainian–Russian mix) and zero otherwise. In both panels, positive post–February 2022 coefficients indicate relatively larger increases in Ukrainian identification and non-Russian replies in rayons with higher prewar 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. Standard errors are clustered at the post-2020 rayon level. Shaded areas indicate 95% confidence intervals.

Language. Columns (3) and (4) of Table 2 examine the use of Russian language in the interview. Panel A 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. Appendix Figure D.6 shows this in the raw data: the difference is almost entirely due to an increase in the use of Ukrainian only, and a comparable decline in Russian. This shift is equivalent to one-fifth of all interviews.

Consistent with our earlier findings, the shift is stronger in rayons with higher prewar shares of ethnic Russians, reinforcing the interpretation of the invasion as inducing national consolidation rather than polarization (Panel B, Columns 3 and 4). Appendix Table D.2 further shows that the adjustment occurs primarily through movement from Russian to Ukrainian, rather than through mixed-language responses. The interaction coefficient is large, positive, and robust to the inclusion of controls in column (4): moving from a rayon with 0% ethnic Russians to one with 25% ethnic Russians in the 2001 Census increases the probability of not replying in Russian by approximately 22 percentage points.

Panel B of Figure 2 shows the event-study estimates. We observe an immediate break after the 2022 invasion; the coefficients remain elevated throughout the postinvasion

period, with no evidence of differential pretrends or subsequent reversal.

4.1.4 Additional Results

Our baseline difference-in-differences specification imposes a linear relationship between identity change and the prewar share of ethnic Russians. To test whether consolidation weakens or reverses in the rayons closest to the Russian prototype, we replace the continuous interaction in Equation (5) with indicators for quartiles of the 2001 rayon-level Russian share. We find no evidence of attenuation: the coefficients increase monotonically across quartiles and are largest in the highest-share quartile (Appendix Table D.3).

We also examine heterogeneity in the effects of the conflict (Appendix Tables D.4–D.7). Younger, urban, and college-educated respondents exhibit larger shifts in ethnic self-identification after 2014 and in language use after 2022, suggesting that these groups lead adjustment across different margins of identity. Older respondents, by contrast, show larger shifts in ethnic self-identification after 2022. Neither outcome exhibits a detectable sex gap in either episode. Since adult outmigration after 2022 was heavily skewed toward women, this suggests a limited scope for selective migration to explain the results.

The two war phases reveal a consistent pattern of national consolidation, but with notable differences. After 2014, adjustment is concentrated in ethnic self-identification, with only modest changes in interview language. After 2022, both outcomes shift sharply toward Ukrainian identity. This contrast is consistent with our framework's Prediction 3: in 2014, the salience and status channels potentially operated in opposing directions, whereas in 2022 the salience shock was stronger and the channels reinforced one another, producing larger identity shifts that extended to the costlier behavioral margin of language use.

4.1.5 Potential Concerns

Four concerns are salient for the above survey-based analysis: social-desirability bias, compositional changes driven by migration and displacement, concurrent language and media policy changes, and ceiling effects.

Social-desirability bias. If respondents became increasingly reluctant to report Russian ethnicity after either conflict episode, one likely manifestation would be increased item nonresponse to the ethnicity question. Appendix Figure D.7 shows no increase in the share

of "no reply" responses after 2014 and a small decline of about 1 percentage point after 2022, from a baseline of 2.3%—the opposite of what increased reluctance would predict and small relative to the shifts in reported ethnicity. Moreover, attitudes toward Russian-speaking Ukrainians deteriorated only modestly after the 2022 invasion (Appendix Figure D.8), making it unlikely that social stigma alone accounts for the much larger shift in interview language. As shown below in Sections 4.3 and 5.1, the post-2022 shift also appears in private behavioral measures, including Google searches and music consumption, which are not observed in the survey and are less susceptible to interviewer demand.[19]

Migration, refugees, and IDPs. Population movements could mechanically change the composition of respondents within locations. We address this concern in four ways. First, our specifications control for displacement: rayon-level IDP-to-population ratios in the 2014 analysis and individual displacement status in 2022. Second, the post-2022 language shift is also evident among military-age men, whose ability to leave Ukraine was generally restricted (Appendix Figure D.9), arguing against selective outmigration as the primary explanation. Third, surveys of Ukrainian refugees in the European Union indicate that approximately 70% report Ukrainian as their primary language of communication (IOM, 2023), compared with 51% of KIIS respondents interviewed in Ukrainian during April 2020–February 2022. To the extent that this difference reflects refugees' greater use of Ukrainian before the invasion, their departure would thus lower the Ukrainian-speaking share among those remaining, attenuating rather than explaining the observed shift. Fourth, the Twitter analysis below is identified entirely from within-account changes, thereby holding user composition fixed; its estimates closely mirror the survey findings.

Concurrent language and media policy changes. A potential concern is the 2019 language law, which expanded the use of Ukrainian in public life (Verkhovna Rada of Ukraine, 2019). Two considerations make it unlikely to explain our main findings. First, its provisions were phased in over several years, whereas our outcomes exhibit a sharp break following the 2022 invasion. Second, the law primarily regulated public and institutional domains and explicitly excluded private communication. As such, respondents remained

[19] Information on KIIS interviewers is limited: we cannot link interviewers across waves, do not observe interviewer characteristics, and only have within-wave interviewer identifiers for 2008–2019. Nevertheless, the ethnic-identification results remain robust to adding interviewer-by-wave fixed effects in the 2014 difference-in-differences analysis (Appendix Table D.8).

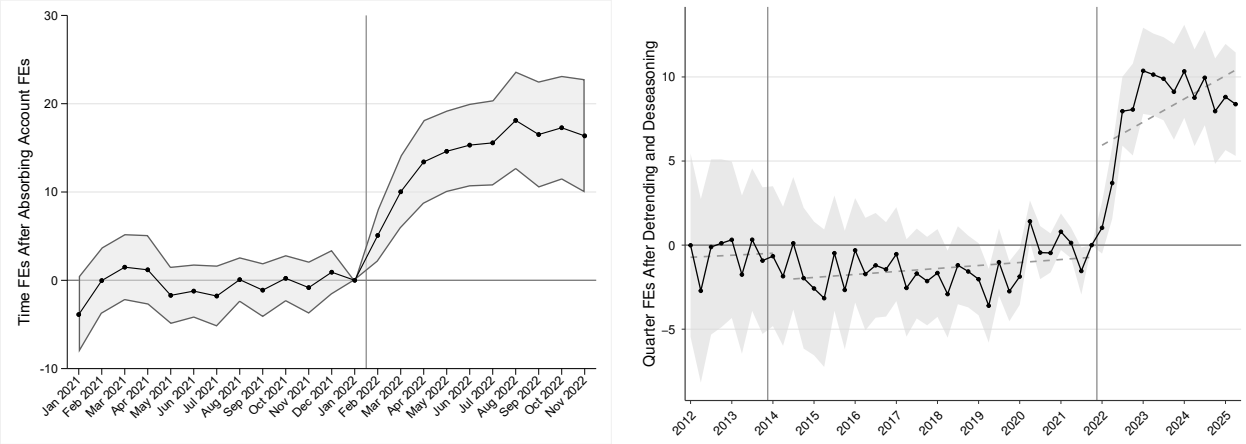
free to choose their KIIS interview language. Similar reasoning applies to the 2017 ban on the Russian social-media platforms VKontakte and Odnoklassniki and the search engine Yandex, as well as the 2021 suspension of pro-Russian television channels. These measures predated the invasion by one to five years, so their direct effects should have appeared earlier or evolved gradually rather than emerged immediately after February 2022. Moreover, as shown below, postinvasion changes extend to behaviors far removed from the scope of these policies, including naming choices. These policies may have contributed to the longer-run expansion of Ukrainian in public life, but they are unlikely to account for the abrupt, broad-based changes observed after the onset of the full-scale invasion.

Ceiling effects. Because Ukrainian ethnic self-identification was already nearly universal in Western Ukraine, low-Russian-share rayons had less scope to shift further, potentially exaggerating the estimated gradient. Three pieces of evidence suggest that ceiling effects do not account for our conclusions. First, the interview language remained far from its ceiling: just before 2022, roughly half of respondents were interviewed in Russian, yet the language outcome exhibits a similar gradient. Second, the gradients remain large and significant among respondents in the Southern and Eastern macroregions and in rayons with an above-median prewar share of ethnic Russians (Appendix Table D.9). Third, and most fundamentally, ceiling effects may contribute to the estimated gradient but do not explain the absence of polarization in its strongest sense: the movement away from Ukrainian identity in rayons with larger Russian populations, as anticipated by Russian forces. Instead, these rayons shift away from Russian and toward Ukrainian identity.

4.2 Language of Tweets

Next, we analyze high-frequency Twitter data, which capture language use by the same individuals over time. We collapse tweets to the account-by-month level and calculate each account's monthly share of Ukrainian-language tweets, which we regress on account and month fixed effects. Account fixed effects isolate within-account changes, while restricting the sample to consistently active accounts limits changes in user composition. We weight accounts equally, preventing highly active users from mechanically driving the estimates.[20]

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



(a) Twitter (within-account, monthly) (b) Google searches (detrended, quarterly)

Figure 3: Revealed Language Use: Twitter Posts and Google Searches

Notes: This figure reports postinvasion shifts toward Ukrainian language in two digital settings. Panel A plots month fixed effects estimated within accounts from Twitter data geolocated to Ukraine between January 2021 and November 2022. The sample includes only accounts that posted in every month from August 2021 through July 2022. For each account-month, we calculate the share of Ukrainian-language tweets among tweets classified as Ukrainian or Russian and regress this share on account and month fixed effects. Estimates are normalized to January 2022, the last preinvasion month; the vertical line marks February 2022. Standard errors are clustered at the account level. Panel B plots quarter fixed effects estimated from Google search data. For each word–oblast–quarter cell, we calculate the share of searches using the Ukrainian rather than the Russian spelling and residualize it with respect to calendar-quarter-by-word-by-oblast fixed effects and pre-2022 word- and oblast-specific linear trends. Dashed lines show fitted linear trends separately for the pre-2014, 2014–2021, and postinvasion periods. Vertical lines mark the onsets of hostilities in 2014 and the full-scale invasion in 2022; 2021Q4 is the omitted reference quarter. The sample excludes Crimea, Sevastopol, and the four oblasts Russia claimed to annex in 2022. Standard errors are two-way clustered by oblast and word. Shaded areas denote 95% confidence intervals.

Panel A of Figure 3 plots the estimated month effects for a balanced panel of accounts active every month from August 2021 through July 2022. The series is flat throughout 2021, but rises sharply after the 2022 invasion. The increase in the share of Ukrainian-language tweets reaches approximately 15 percentage points by June, and remains elevated through the end of the sample. These results closely mirror the survey findings and argue against explanations based solely on interviewer demand or changes in respondent composition. The same users increasingly adopt Ukrainian in public online communication, indicating a persistent shift in linguistic behavior rather than merely in survey reporting.[21]

4.3 Language of Google Searches

Next, we examine private language choice using Google search data. We track Ukrainian and Russian spellings of the same high-frequency words, thereby holding query content

[21]We cannot examine regional heterogeneity in the Twitter data, because the sample of geolocated tweets became increasingly selective by 2022, likely reflecting changes in Twitter's default privacy settings.

fixed and isolating the language in which users search. To account for seasonality and gradual preinvasion Ukrainization, we remove word–oblast-specific seasonal effects and extrapolate word- and oblast-specific linear trends estimated before 2022. Appendix C.2.1 describes the variable construction and estimation procedure.

Panel B of Figure 3 plots deviations from this predicted path. Ukrainian-language searches rise sharply after the 2022 invasion. By the third quarter of 2022, the Ukrainian share is 8 percentage points above its predicted preinvasion path and remains 8–10 percentage points higher through the end of the sample. Together with the survey and Twitter results, this finding reveals a large and persistent linguistic shift that extends to private behavior and is less susceptible to interviewer demand or reputational concerns. Its smaller magnitude relative to the 20-percentage-point shift in interview language is consistent with social-image concerns amplifying the survey response, although differences in populations and behaviors may also contribute. The lack of a language shift following the 2014 war phase is also consistent with the survey estimates.

Appendix Figure D.10 examines heterogeneity by prewar ethnic composition. We interact quarter indicators with the 2001 oblast-level share of ethnic Russians, taking 2021Q4 as the reference period. The interaction coefficients are close to zero and exhibit no systematic trend before the invasion, and then turn positive thereafter. Thus, oblasts with larger prewar Russian populations experience larger shifts toward Ukrainian-language searches, reinforcing the earlier survey-based evidence of national consolidation.

5 Results: Cultural Shifts

Culture is central to theories of nation-building (Anderson, 1983; Gellner, 1983). In this section, we examine whether the shift toward Ukrainian identity extends to cultural behavior. Our measures span domains that differ in visibility and adjustment cost: music consumption and searches for food and recipes capture routine—and largely private—behavior, while baby names are high-stakes, long-term, publicly observable household choices.[22] Together, they test Prediction 2 of our framework.

[22]See, for example, Atkin et al. (2021) on using food consumption to study identity choice.

5.1 Popular Music and Spotify Charts

Music listening is a widespread, routine, largely private form of cultural consumption. It provides a revealed-preference measure of identity that is both naturally occurring and less susceptible to social-desirability concerns than survey responses. The music market is particularly informative: it allows us to observe changes both along the demand side (listeners) and the supply side (recording artists).

Chart-level language dynamics. Spotify entered Ukraine in July 2020 and quickly became the country's leading streaming platform, reaching a 45% market share by 2023 (Statista, 2025). We document changes in the language composition of songs appearing in Spotify's weekly charts of the top-200 songs in Ukraine.

Figure 4 reports the share of chart entries in each language group (Ukrainian, Russian, and Other) and shows a massive decline in the popularity of Russian-language music after the full-scale invasion. Before February 24, 2022, Russian-language songs accounted for roughly 60% to 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 to almost 25%. Appendix Figure D.11 shows that this reallocation also appears in absolute streaming volumes: weekly streams of Ukrainian-language charting songs rose from very low preinvasion levels to several million, while Russian-language streams fell sharply. The shift continued thereafter: by early 2023, Ukrainian-language songs had overtaken Russian-language songs in chart presence, with partial substitution toward other languages (primarily English).[23]

Demand-side reallocation among preexisting songs. To assess whether the postinvasion shift in listening patterns reflects changes in listener demand rather than shifts in the supply of new music, we repeat the analysis using only songs released before February 24, 2022 (Appendix Figure D.12). We observe the same pattern with similarly large magnitudes: listening to Russian-language songs declines sharply after the invasion, while Ukrainian- and other-language songs gain substantially in popularity.

[23]The chart share of songs with missing lyrics remains relatively stable and generally below 10% throughout. Lyrics may be missing because a track is instrumental or they have not yet been uploaded to Genius.

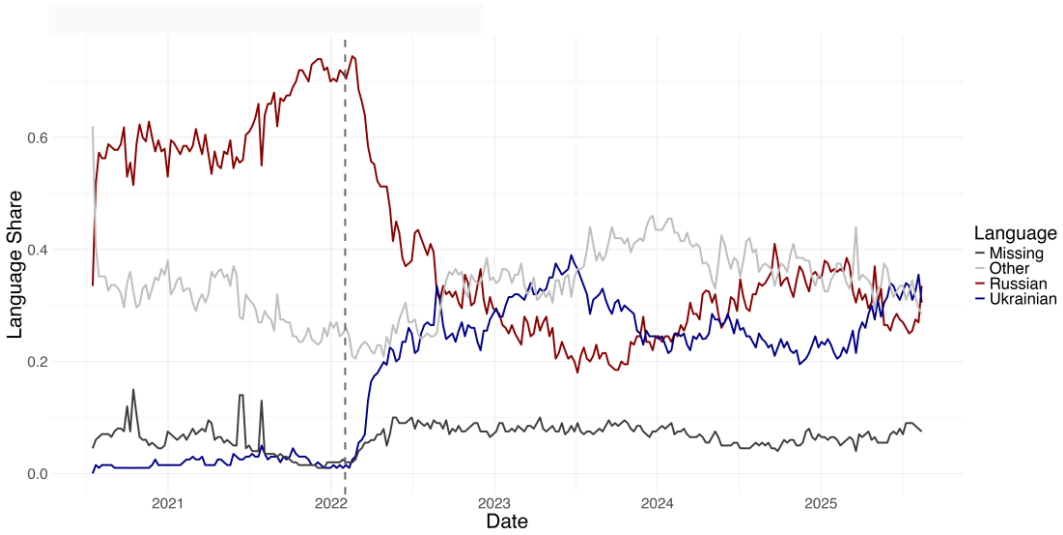


Figure 4: Language Composition of Spotify's Weekly Top Ukraine Charts, 2020–2025

Notes: This figure reports the weekly share of tracks in Spotify's weekly Top 200 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 grouped under Missing.

Language preferences and artist boycotts. An interesting question is whether the postinvasion shift in music consumption reflects changing language-linked identity preferences, or a reallocation away from particular artists, possibly reflecting in-group bias. To address this question, we collect artist-level data, including data on their stance toward the war where available (see Appendix C.4 for details). We obtain three findings. First, to separate song language from artist origin, we restrict the sample only to Ukrainian artists. We find a similarly sharp reversal within this subsample (Panel A of Appendix Figure D.13). The share of Russian-language songs falls from roughly 75% to below 20% within a month of the invasion, with a corresponding rise in the popularity of Ukrainian-language songs.

Second, we further restrict the sample to *prewar releases* by Ukrainian artists. Even among these songs we observe a sharp spike in the popularity of songs in Ukrainian following the invasion, and a drop in the share of Russian-language songs in the top charts (Panel B of Appendix Figure D.13). However, in this case the gap attenuates over time, as the popularity of prewar Russian-language songs by Ukrainian artists partly recovers in 2024–2025. Old favorites by in-group artists may be hard to let go of.

Third, we examine the change in popularity of songs by *Russian* artists whose stance toward the war was coded by the Stars About the War Project as either prowar or antiwar.

We find similar declines in the first two years after the invasion, regardless of whether the artists support or oppose the aggression (Appendix Figure D.14). Although the popularity of antiwar Russian artists appears to recover somewhat in 2024–2025, the differences between the two series are statistically indistinguishable from zero. Together, these results indicate that the reallocation is not driven solely by boycotts based on artist origin or political stance, but also reflects a broader shift in language-based listening preferences.

Supply response. We examine language choice in new releases by Ukrainian artists. Before 2022, roughly 60% of quarterly releases were in Russian and 20% in Ukrainian (Panel A of Appendix Figure D.15). By the second quarter of 2022, Ukrainian-language releases had overtaken Russian-language releases. By mid-2023, the prewar pattern had nearly reversed, with close to 60% of new songs recorded in Ukrainian and fewer than 30% in Russian. The shift is also present among artists active before the invasion (Panel B of Appendix Figure D.15) and even more pronounced among artists first observed after the invasion: among these artists 80% to 90% of new releases are in Ukrainian and fewer than 5% are in Russian (Panel C of Appendix Figure D.15).

To distinguish within-artist switching from changes in artist composition, we regress each artist's quarterly language shares on quarter indicators and artist fixed effects.[24] For both languages, the quarterly indicators are mostly insignificant prior to the invasion (Appendix Figure D.16). This is followed, however, by a sharp, roughly 20-percentage-point increase in Ukrainian-language releases after the invasion, and a comparable decline in Russian-language releases within artists. Taken together with the previous results, the supply-side reorientation thus reflects both entry and substantial within-artist switching. These patterns may reflect artists' responses to changing listener demand, changes in their own identities, or both (our data do not allow us to distinguish between these channels).

5.1.1 Potential Concerns

Coverage in occupied regions. A potential concern is that war-related disruptions could have changed the regional composition of Spotify users and mechanically contributed to the observed language shift. Several facts argue against this explanation. Donetsk and

[24] For each language l, we estimate $share^{l}_{at} = \theta^{l}_{t} + \mu_a + \varepsilon^{l}_{at}$, where $share^{l}_{at}$ denotes the share of releases by artist a in quarter t in language l, θ^{l}_{t} are language-specific quarter fixed effects, and μ_a are artist fixed effects. Standard errors are clustered at the artist level.

Luhansk oblasts already had the lowest Spotify-related Google search interest before the full-scale invasion (Appendix Table D.10), while territories occupied after 2022 accounted for at most 9.6% of the preinvasion population under Ukrainian control.[25] Moreover, the results reflect a sharp increase in the absolute volume of Ukrainian-language streams (Appendix Figure D.11). As discussed earlier, surveys indicate that Ukrainian refugees have been disproportionately Ukrainian-speaking, implying that selective emigration would, if anything, attenuate rather than generate the observed reallocation.

Recommendations and platform promotion. Another potential concern is that Spotify may have amplified the rise of Ukrainian-language music through curated playlists, artist promotion, or its recommendation system. Spotify undertook several visible initiatives supporting Ukrainian music after the invasion (Spotify Newsroom, 2022, 2023; Suspilne Media, 2024), so some amplification cannot be excluded. Several facts, however, make platform promotion unlikely to account for the shift on its own. First, the reallocation also appears among preinvasion releases and within Ukrainian artists, rather than only among newly promoted content, and parallel language changes occur outside Spotify, in survey responses, Twitter use, and Google searches. Second, recommendation systems that adapt to aggregate listening behavior would reinforce an underlying change in demand rather than constitute an independent mechanical confound. Third, Spotify's removal of several prowar Russian artists occurred only in June 2024, and only one song by the affected artists ever entered Ukraine's Top-200 (Radio Free Europe/Radio Liberty, 2024). Platform promotion may therefore have amplified the reallocation, but is unlikely to have generated its timing, magnitude, or cross-domain character.

Public music consumption. Another potential concern is that the rise of Ukrainian-language music may reflect changes in music played in public venues rather than a broad shift in listeners' private consumption preferences. This explanation would be especially plausible if the aggregate pattern were concentrated in dance and electronic music, both of which are frequently consumed in public settings. Using artist metadata from the Last.fm API, we classify songs into six broad genres. Appendix Figure D.17 shows that the

[25]Subsequently occupied territories contained approximately 3.59 million of the 37.44 million residents living in areas controlled by Ukraine as of January 1, 2022 (Ukrstat, 2022). This calculation does not account for the large outflow from war zones, estimated at 2.8 million displaced persons (Interfax-Ukraine, 2024), which would reduce the number of possible affected listeners even further.

Ukrainian-language share rises within every genre, including those less associated with public-venue listening, while the Russian-language share declines throughout. Overall genre composition remains largely stable, apart from some reallocation from hip-hop and rap toward pop. These patterns make it unlikely that the aggregate language shift is driven by public-venue listening or changes in genre composition.

5.2 Baby Names

Names provide an especially durable measure of identity: they are publicly observable, persistent, and costly to reverse (Fouka, 2019; Esposito et al., 2023).

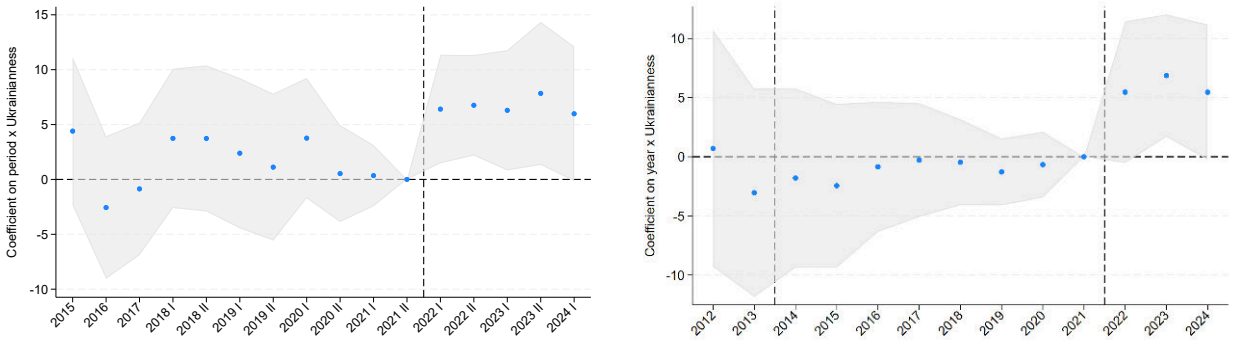
To examine whether baby naming shifted toward more-Ukrainian names after Russia's full-scale invasion, we assign each name a time-invariant Ukrainianness score based on its pre-2022 relative popularity in Ukraine and Russia (as described in Appendix C.3). For example, Bohdan and Solomiya receive relatively high Ukrainianness scores and Yelysei and Yeseniia relatively low ones. We then estimate how the association between this score and a name's probability of appearing in the list of most popular baby names by oblast, published by the Ukrainian Ministry of Justice, evolves over time. Specifically, we estimate

$$\text{Popular}_{nto} = \sum_{s \neq 2021\text{ II}} \theta_s \left(\text{Ukrainianness}_n \times \mathbb{1}[t = s]\right) + \eta_{no} + \mu_{to} + \varepsilon_{nto}, \tag{6}$$

where Popular_{nto} indicates whether name n appears on the popularity list for oblast o in period t, and we include oblast–name fixed effects η_{no}, so that identification comes from within-oblast changes in name popularity. Oblast–time fixed effects μ_{to} absorb variation in list length and other oblast-level shocks. The omitted period is the second half of 2021. Standard errors are two-way clustered by oblast and name.

Panel A of Figure 5 shows the estimated coefficients θ_s. We observe no systematic pretrends, followed by a clear increase in the relative popularity of more-Ukrainian names after the invasion.[26] The corresponding pre–post estimate (Appendix Table D.11) implies that the postinvasion increase in the probability of appearing on a regional popularity list is 5.1 percentage points larger for a name with a Ukrainianness score of one than for a name with a score of zero, equivalent to approximately 38% of the sample mean. Estimates are positive for both boys' and girls' names, although somewhat larger and more precisely

[26] Results are robust to an alternative specification using the average Ukrainianness of listed names at the oblast-by-period level (Appendix Figure C.3).



(a) Popularity of Ukrainian baby names

(b) Popularity of Ukrainian cuisine

Figure 5: Cultural Reallocation Toward Ukrainian Prototype: Baby Names and Cuisine

Notes: Each panel reports event-study coefficients that trace how the association between a cultural marker's prewar Ukrainianness and its popularity evolves over time. In both panels, coefficients are on interactions between period indicators and a time-invariant Ukrainianness score constructed from prewar relative popularity in Ukraine vs. Russia; positive post-2022 coefficients indicate that more-Ukrainian names (in Panel A) or dishes (in Panel B) become relatively more popular after the full-scale invasion. Panel A estimates Equation (6) at the name-oblast-period level, where periods are years before 2018 and half-years thereafter; the outcome is an indicator equal to one if name n appears in the Ministry of Justice list of most popular names in oblast o in period t; the specification includes oblast–period and oblast–name fixed effects; coefficients are normalized to the second half of 2021; standard errors are two-way clustered by name and oblast. Panel B estimates Equation (7) at the dish-oblast-month level, with annual coefficients; the outcome is the raw Google Trends score for dish i in oblast o and month t; the specification includes dish-by-oblast and oblast-by-month fixed effects; coefficients are normalized to 2021, the omitted reference year; 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 in both panels.

estimated for girls. Appendix Figure C.4 illustrates these changes for individual names, with more-Ukrainian names gaining popularity on average.

Appendix Table D.11 further clarifies the nature of this shift. The postinvasion increase in the relative popularity of Ukrainian names is concentrated in oblasts with above-median popularity-list lengths (column 5). This pattern is consistent with Ukrainian names crossing publication thresholds farther down the name distribution, suggesting that the reallocation extended beyond the most popular names. We also find suggestive evidence of larger increases in oblasts with higher prewar shares of ethnic Russians. Although this heterogeneity is imprecisely estimated ($p = 0.156$; column 6), its direction is consistent with the patterns of national consolidation above.

As an external validation, we examine administrative birth records from Kazakhstan, which report the universe of names given to children by recorded ethnicity. Applying the same prewar Ukrainianness scores to names given to ethnic Ukrainian children, we find a similar post-2022 shift toward more-Ukrainian names (Appendix Figure C.5). Ethnically

Ukrainian parents in Kazakhstan are less directly exposed to the institutional and economic incentives that could make Ukrainian-coded names advantageous within Ukraine. The parallel shift outside Ukraine therefore provides evidence that the naming response is not driven solely by strategic adaptation to the domestic wartime environment and is consistent with a broader increase in the relative appeal or status of Ukrainian identity.

5.3 Cuisine

Culinary interests are an everyday and largely private dimension of cultural behavior. We use Google searches for foods and recipes as a high-frequency behavioral measure of interest in national cuisine.[27] For each dish, we construct a time-invariant Ukrainianness score based on its preinvasion relative search popularity in Ukraine and Russia and estimate how the relative popularity of more- and less-Ukrainian dishes evolves over time. Appendix C.2.2 describes the data and score construction.

We again estimate an event-study specification that allows the association between dish Ukrainianness and search popularity to vary by year, controlling for dish-by-oblast and oblast-by-month fixed effects.[28]

Panel B of Figure 5 reports the yearly event-study estimates. The estimates are stable before 2022 and rise thereafter, indicating greater relative search intensity for more-Ukrainian dishes within the same oblast-month. Appendix Figure D.18 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 increase is approximately 0.2–0.3 standard deviations.

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

[27] Online recipe use is widespread in modern settings: e.g., in a 2024 survey of U.S. consumers, 91% reported using online recipes (Chicory, 2024).

[28] Formally, we estimate

$$\text{Popularity}_{iot} = \sum_{s \neq 2021} \theta_s \left(\text{Ukrainianness}_i \times \mathbb{1}[\text{year}(t) = s]\right) + \mu_{io} + \delta_{ot} + \varepsilon_{iot} \tag{7}$$

where Popularity_{iot} denotes Google Trends search intensity for dish i in oblast o in month t. The baseline outcome is the raw Google Trends score, normalized by its within-oblast maximum over 2012–2024; an alternative outcome is standardized to have a mean of zero and unit variance within dish. The fixed effects μ_{io} absorb time-invariant differences in dish popularity across oblasts; δ_{ot} absorb common shocks to search activity within oblast-month. The coefficients θ_s measure how the association between dish Ukrainianness and its popularity differs relative to 2021, the omitted reference year. Standard errors are two-way clustered by dish and oblast.

unlikely to be driven by the disappearance of Russian brands or by reduced availability of imported products from Russia.

To sum up this section, across all three cultural domains, the evidence points in the same direction. Music, cuisine, and baby names vary in visibility, adjustment costs, and time horizon, yet they all exhibit a clear reallocation toward Ukrainian cultural markers, consistent with Prediction 2. The music evidence also suggests a self-reinforcing dynamic: as listeners shifted toward Ukrainian-language music, existing artists reoriented their output and new entrants released overwhelmingly in Ukrainian, expanding the available repertoire and potentially lowering the future cost of tilting one's consumption toward the national prototype. More broadly, these patterns illustrate how identity change can reshape the cultural environment in ways that can reinforce it.

6 Channels of Consolidation and the Scope for Polarization

Having documented broad national consolidation, we now ask whether the shift primarily reflects rejection of Russia, or stronger attraction to the Ukrainian nation, or both; whether direct exposure to violence amplifies the shift; and whether consolidation reverses under the very different military and political conditions prevailing in the occupied territories following the 2014 annexation of Crimea and the war in Donbas.

6.1 Out-Group Rejection

National consolidation may arise through rejection of the external out-group, greater attraction to the national in-group as its status and perceived proximity increase, or both. In the case of Ukraine, out-group rejection is clearly present. Attitudes toward Russians deteriorate after both episodes (Appendix Figure D.19): the shift is limited after 2014 but dramatic after 2022, when the share of respondents unwilling to admit Russians living in Russia into Ukraine rises from about 25% to roughly three-quarters.

Rejection alone, however, cannot fully explain the consolidation we observe. After both episodes, the rise in Ukrainian ethnic self-identification is accompanied by a decline not only in Russian identification but also in reporting other ethnicities such as Belarusian and Jewish (Appendix Figures D.5 and D.20). Because these respondents had no Russian identity to disavow, this pattern cannot be explained solely by its stigmatization. Moreover, the

hostility that emerges is directed primarily toward the external enemy: after 2022, attitudes toward Russians living in Russia deteriorated far more than attitudes toward Russians living in Ukraine or Russian-speaking Ukrainians (Appendix Figures D.19, Panels B–C, and D.8). The salient boundary therefore appears to be Ukraine vs. Russia as the aggressor, rather than the internal divide between Ukrainian- and Russian-speaking citizens.

6.2 Local Exposure to Violence

Is local exposure to violence, possibly the most acute form of salience shock in our framework, associated with stronger movement toward a Ukrainian identity? We link the KIIS data to the VIINA database (Zhukov, 2022), which records the location and timing of air, artillery, and drone attacks in Ukraine; civilian- and military-casualty events; and incidents involving hospitals. For each rayon and survey wave, we construct per capita counts of events occurring in that rayon since the preceding wave.

We examine whether respondents become less likely to reply in Russian following greater local exposure to violence within rayons, while accounting for nationwide shocks and differential language dynamics by prewar ethnic composition. We exclude displaced respondents, for whom events in their home rayon may not reflect local exposure.[29]

Appendix Table D.12 reports estimates for all nonoccupied rayons and for a stricter sample of never-occupied rayons farther from the front. Across the four event measures and both samples, every coefficient is negative and all but one are statistically significant: rayons experiencing more violence exhibit larger declines in Russian-language replies. In the nonoccupied sample, a one-standard-deviation increase in exposure to air, artillery, and drone attacks between waves is associated with a 1.1-percentage-point lower probability of replying in Russian, about 4% of the sample mean. Because exposure to violence is not randomly assigned, we interpret these estimates as corroborative rather than as separate causal effects. Their consistency nevertheless supports the salience channel in our frame-

[29] Specifically, for each event type, we estimate

$$Y_{irt} = \beta T_{rt} + X'_{irt}\delta + \mu_r + \lambda_t + \varepsilon_{irt} \tag{8}$$

where Y_{irt} indicates whether respondent i from preinvasion rayon r replied in Russian in wave t, and T_{rt} is the event count per capita in rayon r between waves $t-1$ and t. Rayon and wave fixed effects absorb persistent local differences and nationwide shocks. X_{irt} includes respondent demographics and the rayon's prewar share of ethnic Russians interacted with wave indicators, allowing for differential language dynamics by ethnic composition that may correlate with the incidence of attacks (Gorbuntsova et al., 2026). Standard errors are clustered by rayon.

work: conditional on nationwide shocks, persistent rayon differences, and differential wave dynamics by prewar ethnic composition, greater local exposure to Russia's attacks is associated with a sharper movement away from Russian language.

6.3 Identity in Occupied Territories

Our framework implies that consolidation is not automatic: whether external conflict consolidates or polarizes national identity depends on local military and political conditions. The framework is symmetric in this respect. Reversing the direction in which attention is reallocated reverses the identity response (Appendix B.4), so where occupation made the *within*-Ukraine boundary dominant rather than the interstate one, the same mechanism that produces consolidation elsewhere predicts divergence. The 2014 events may capture this contingency, as they differed from 2022 on two margins. The status shock likely had the same sign across Ukraine: Russia annexed Crimea and established control over parts of Donbas rapidly, against a limited international response. This plausibly raised the relative status of Russian identity, possibly more in the areas Russia actually occupied. In contrast, the salience shock may have differed in sign across areas. Elsewhere in Ukraine, Russia's aggression was perceived as an attack on the country as a whole, sharpening the *Ukraine vs. Russia* divide and contributing to consolidation. Within the occupied territories, however, things may have been different. Russia's swift annexation of Crimea met little armed resistance in a population that was, on average, closer to the Russian prototype than the rest of Ukraine.[30] Russia also took pains to disguise its involvement and cast the conflict as an *internal* dispute with grassroots support. Crimea was taken by unmarked armed forces, and the Donbas "people's republics" were presented as local secessions. Thus, the shock may have heightened the internal ethnolinguistic divide.

Consistent with this possibility, in Crimea Google searches in Ukrainian fell from an already low preannexation share of roughly 3%–4% to approximately 1%, the opposite of the subsequent increase in government-controlled Ukraine (Appendix Figure D.21). Census figures likewise suggest a decline in Ukrainian identification, although evidence collected under occupation is difficult to interpret.[31]

[30] In the 2001 Census, ethnic Russians constituted 58% of Crimea's population and 71% of Sevastopol's, while 79% of residents reported Russian and 9.5% Ukrainian as their mother tongue. In KIIS surveys in 2008–2014, only 1.6% of respondents were interviewed in Ukrainian or a Ukrainian–Russian mix.

[31] The share identifying as Ukrainian fell from roughly 24% in the 2001 Ukrainian Census to 7.5% in

Occupied Donbas provides more-direct survey evidence. KIIS continued surveying residents of the areas under Russian control through 2015. In Appendix Table D.13, we augment the baseline specifications with an indicator for rayons of Donetsk and Luhansk oblasts that came under Russian control after February 2014 and its interaction with the post-2014 period. This interaction is large and negative for Ukrainian ethnic identification (Panels A and B). The reversal is particularly striking because these rayons had high prewar shares of ethnic Russians, a characteristic associated with stronger consolidation elsewhere in Ukraine; controlling for the ethnic-Russian-share gradient leaves the occupation interaction, if anything, larger. Interview language exhibits no significant differential change, consistent with the muted aggregate language response after 2014.

These estimates require caution. KIIS ceased surveying non-government-controlled territories after 2015, the remaining samples are small and potentially affected by intimidation, and large-scale displacement altered the composition of the surveyed population. Subject to these caveats, Crimea and occupied Donbas provide a suggestive counterpoint to the main results and further validation of our framework. National consolidation is not an automatic consequence of external war: status and salience forces can produce movement toward either identity depending on local tactics and military outcomes.

7 Ukrainian Identity and Attitudes Toward Concessions

National cohesion is often viewed as a foundation of collective defense: a stronger common identity helps societies sustain mobilization and resist external aggression (Alesina et al., 2020; Qian and Tabellini, 2026). Although our data do not measure participation in the war, they can capture a related dimension of societal resolve: willingness to accept territorial concessions in exchange for peace. If territory is seen as integral to the nation, stronger Ukrainian identification should be associated with greater opposition to ceding it. Using KIIS surveys fielded between May 2022 and February 2025, we relate support for territorial concessions to interview language and self-reported Ukrainian ethnicity.[32]

the 2021 Russian census. These figures may reflect reporting pressures, measurement problems, and compositional change. Available estimates suggest that 20,000–50,000 residents left Crimea after annexation, equivalent to roughly 1% to 2% of its population.

[32]We estimate the following: $\text{Concession}_{iot} = \beta\,\text{Identity}_{iot} + X'_{iot}\delta + \mu_o + \gamma_t + \varepsilon_{iot}$, where Concession_{iot} indicates support for territorial concessions by respondent i from preinvasion oblast o in survey wave t. Identity is measured by interview language—Ukrainian or a Ukrainian–Russian mix, relative to Russian—and by self-reported Ukrainian ethnicity. X_{iot} contains demographic and displacement controls.

Table 3: Ukrainian Identity and Readiness for Territorial Concessions

	Willingness to concede territory for peace (=1)				
	(1)	(2)	(3)	(4)	(5)
Sample	All	All	All	All	Men $<$ 60 y.o.
Panel A: Interview language (ref. = Russian)					
Ukrainian language	-0.108***	-0.148***	-0.150***	-0.146***	-0.145***
	(0.010)	(0.010)	(0.010)	(0.019)	(0.018)
Mixed language	-0.120***	-0.121***	-0.123***	-0.089***	-0.151***
	(0.013)	(0.013)	(0.013)	(0.024)	(0.022)
Dep. var. mean	0.193	0.193	0.193	0.195	0.202
Dep. var. std. dev.	0.395	0.394	0.394	0.397	0.402
Observations	15,872	15,814	15,814	6,161	5,019
Panel B: Self-reported ethnicity					
Ukrainian ethnicity	-0.181***	-0.188***	-0.188***	-0.173***	-0.208***
	(0.018)	(0.017)	(0.017)	(0.031)	(0.031)
Dep. var. mean	0.187	0.187	0.187	0.189	0.190
Dep. var. std. dev.	0.390	0.390	0.390	0.392	0.392
Observations	15,750	15,712	15,712	6,079	5,039
Controls		✓	✓		✓
Preinvasion oblast FE		✓			
Survey-wave FE		✓			
Oblast $\times$ wave FE			✓		✓
Oblast $\times$ wave $\times$ controls FE				✓	
Sample: males aged $<$60					✓

Notes: This table reports OLS estimates of the association between Ukrainian identity and willingness to cede territory in exchange for peace. The outcome equals one for respondents who report readiness for territorial concessions and zero for those who do not. Panel A measures identity through the language of the interview (Ukrainian or a Ukrainian–Russian mix relative to Russian as the reference category), while Panel B uses an indicator for self-reported Ukrainian ethnicity. Columns (1)–(4) use the full analysis sample, while column (5) restricts the sample to males under 60 (the population most directly exposed to military mobilization). Column (2) includes preinvasion-oblast and survey-wave fixed effects; column (3) replaces these with their full interaction (oblast $\times$ wave fixed effects); column (4) is a matching estimator that absorbs fixed effects for each cell defined by the full interaction of preinvasion oblast, survey wave, sex, age group, education, settlement type, and displacement status; column (5) repeats column (3) on the male-under-60 subsample. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 reports increasingly saturated specifications, culminating in comparisons among respondents from the same preinvasion oblast and survey wave with the same demographic and displacement characteristics. We also report estimates for men under 60, for whom the personal stakes of continued resistance are likely to be especially high.

Across all specifications, expressed Ukrainian identity is strongly negatively associated with support for territorial concessions. In the most saturated specification, in which respondents are matched on a range of observables, respondents interviewed in Ukrainian are 14.6 percentage points less likely to support concessions than Russian-language respondents, while those interviewed in a Ukrainian–Russian mix are 8.9 percentage points

less likely to support them. Self-reported Ukrainian ethnicity is also associated with a 17.3-percentage-point lower probability of supporting concessions, relative to a sample mean of 18.9%. The associations remain similar among men under 60.[33]

These correlations are descriptive: identity and wartime attitudes may be jointly shaped by unobserved experiences, and the direction of causality cannot be established. Nonetheless, their stability across specifications suggests that observable geographic and demographic differences account for little of the association. Overall, these estimates suggest that the national consolidation documented above may be accompanied by increased resilience and substantially lower willingness to accept territorial compromise.

8 Conclusion

This paper studies how large-scale external conflict reshapes national identity in a historically heterogeneous society with groups that have close ties to the aggressor. 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 administrative records and digital traces of public and private behavior, allowing us to track identity change across domains that differ sharply in visibility, persistence, and adjustment cost.

We find evidence of national consolidation rather than polarization. Both episodes produce large and persistent shifts toward Ukrainian ethnic identification, while linguistic adjustment is modest after 2014 but sharp after 2022. In both episodes, movement toward Ukrainian identity is largest in districts with higher prewar shares of ethnic Russians. 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 cultural and linguistic ties to the aggressor.

The changes extend well beyond stated identity. Following the 2022 invasion, language use shifts sharply toward Ukrainian not only in survey interviews but also within Twitter accounts and in private Google searches. Cultural behavior moves in parallel: listening shifts toward Ukrainian-language music, searches toward more-Ukrainian cuisine, and

[33] A coarser matching estimator that absorbs oblast × wave × sex × age-group cells and controls for education, settlement type, and displacement retains 97% of the sample and yields similar estimates. Applying the column 4 matching estimator to men under 60 also produces similar results.

baby-naming choices toward more-Ukrainian names (the last also among ethnic Ukrainians in Kazakhstan). These outcomes span margins with very different visibility, adjustment costs, and time horizons, yet all move in the same direction. Their joint movement supports an interpretation of broad, behaviorally meaningful identity change rather than a shift confined to survey reporting or public expression.

The evidence is consistent with an endogenous-identity framework in which conflict reshapes identification through its effects on group status, the salience of group boundaries, and distance from competing national prototypes. In 2022, all three forces operated in the direction of consolidation. Ukraine's effective resistance and international support plausibly raised the relative status of Ukrainian identity, while the breadth of the invasion sharpened the Ukraine–Russia boundary; behavioral adjustment may then have reinforced the initial shift by reducing intranational differences. In 2014, by contrast, Russian rapid gains and covert strategy likely offset part of the salience effect, yielding a more muted response along the costlier margin of language use. Suggestive evidence from Crimea and occupied Donbas further underscores that consolidation is contingent rather than automatic: where Russia rapidly established control, identity seems to have moved in the opposite direction.

Overall, the Ukrainian case highlights the malleability of national identity under external threat. Identity is often treated as slow-moving and deeply rooted, yet our evidence shows that armed conflict can generate rapid, persistent, and behaviorally meaningful change even where fragmentation once appeared plausible. These findings matter for theories of nation-building, for our understanding of collective resilience during wartime, and for broader debates about the political and economic consequences of war.

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.

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*.

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

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, December). Nation-building, nationalism, and wars. *Journal of Economic Growth 25*(4), 381–430.

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*.

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

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*(4), 1193–1251.

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 (2026). Religion, Education, and the State. *Review of Economic Studies 93*(3), 1494–1535.

BBC News Russian (2022). Zelensky in Munich: There Are Better Ways to Help Ukraine Than Warnings of a Russian Invasion. BBC Russian Service.

Boulègue, M., J. Bronk, N. Bugayova, M. Clark, R. Häggblom, M. Konaev, and P. Luzin (2023). Understanding Russia's Military Vulnerabilities. *Chatham House*.

Bureau of National Statistics of Kazakhstan (2026). Popular names. `https://stat.gov.kz/ru/instuments/name/`. Accessed August 2026.

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

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.

Chicory (2024). *The State of Online Recipes*. Chicory.

Coser, L. (1956). The functions of social conflict.

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

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.

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.

Egorov, K., V. Korovkin, A. Makarin, and D. Nigmatulina (2026). Export Sanctions: Rerouting, Disruption, and the Reach of Economic Warfare. Technical report.

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 (2026). The Price of War. *American Economic Review 116*(3), 791–827.

Federle, J., A. Meier, G. J. Müller, and V. Sehn (2022). Proximity to War: The Stock Market Response to the Russian Invasion of Ukraine. *Journal of Money, Credit and Banking*.

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.

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

Genicot, G. and D. Ray (2024). Identities in conflict. In *Handbook of the Economics of Conflict*, Volume 1, pp. 201–247. Elsevier.

Gidron, N. and M. Shayo (2026). The political economy of national identity. In G. A. Akerlof and R. E. Kranton (Eds.), *Handbook of the Economics of Identity*. Forthcoming.

Gorbuntsova, A., G. Khanna, and S. Mehmood (2026). The Domestic Political Economy of War: Evidence from Russia. Working Paper 35107.

Hauk, E. (2025). The effect of foreign influence on conflict and social identity in ethnically diverse societies. *Games 16*(6), 65.

Interfax-Ukraine (2024, July). More than 2.8 Million Civilians Evacuated from Dangerous Territories.

IOM (2023). Ukrainian Refugee Response in the European Union: Needs, Intentions, and Integration Challenges. Technical Report 74.

Jha, S. (2013). Trade, Institutions, and Ethnic Tolerance: 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.

Jones, S. G. and R. McCabe (2026). Russia's Grinding War in Ukraine. *Center for Strategic and International Studies*.

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 (2026). Supply Chain Disruption and Reorganization: Theory and Evidence from Ukraine's War. *Review of Economic Studies 93*(4), 2750–2783.

Kravtsiv, B. and V. Kubijovyc (1993). Russification. In *Encyclopedia of Ukraine*.

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.

Le Monde (2024). Ukrainian as a language of resistance: "When I hear Russian, it makes me want to vomit".

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

Osypenko, O. (2025). Wartime resurgence: Surzhyk, a blend of Ukrainian and Russian once stigmatised as a marker of backwardness, is enjoying an unexpected comeback.

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

Pop-Eleches, G. and G. Robertson (2018). Revolutions in Ukraine: Shaping Civic Rather Than Ethnic Identities. PONARS Eurasia Policy Memo No. 510.

Qian, N. and M. Tabellini (2026). Racial discrimination and the social contract: Evidence from US Army enlistment during WWII. *Review of Economic Studies 93*(2), 1261–1295.

Radio Free Europe/Radio Liberty (2024, June). Spotify Removes Russian Artists Who Support Ukraine War.

Rapawy, S. (1997). Ethnic Reidentification in Ukraine. IPC Staff Paper 90.

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

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. (2020). Social identity and economic policy. *Annual Review of Economics 12*, 355–389.

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

Shayo, M. and A. Zussman (2017). Conflict and the persistence of ethnic bias. *American Economic Journal: Applied Economics 9*(4), 137–165.

Silva, F. B. G. and E. Volkova (2018). Can VPIN forecast geopolitical events? Evidence from the 2014 Crimean Crisis. *Annals of Finance 14*(1), 125–141.

Simmel, G. (1904). The Sociology of Conflict. I. *American Journal of Sociology 9*(4), 490–525.

Spotify Newsroom (2022). Spotify Mobilizes Fans, Artists, and Our "Music for Ukraine" Playlist Ahead of Ukrainian Independence Day.

Spotify Newsroom (2023). Jamala Kicks Off Spotify EQUAL's Ukrainian Launch.

StatCounter (2021). Search Engine Market Share in Ukraine.

Statista (2025). Digital music — ukraine: Market data and outlook.

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

Suspilne Media (2024). Kalush orchestra released its second project for spotify singles.

The UN Refugee Agency (2025). Ukraine emergency response.

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

Treisman, D. (2018). *The New Autocracy: Information, Politics, and Policy in Putin's Russia*.

Ukrstat (2001). All-Ukrainian Population Census 2001: Linguistic Composition of the Population.

Ukrstat (2022). Population estimates as of january 1, 2022.

Verkhovna Rada of Ukraine (2019). On Ensuring the Functioning of the Ukrainian Language as the State Language. Adopted April 25, 2019.

Weigand, M., C. Mohr, and D. Cantoni (2025). Identity and Institutional Change: Evidence from First Names in Germany, 1700–1850. Technical report, CESifo Working Paper.

Zhukov, Y. M. (2022). VIINA: Violent Incident Information from News Articles on the 2022 Russia–Ukraine War. GitHub repository.

Online Appendix

War and Nation-Building: Evidence from Ukraine

Serhii Abramenko Vasily Korovkin Alexey Makarin Moses Shayo

Contents

A Literature Review **2**

B Conceptual Framework: Details **8**

B.1 Setup 8

B.2 Identity Choice 9

B.3 Behavioral Adjustment 10

B.4 Comparative Statics 11

B.5 Summary 18

C Additional Details on Data Construction **21**

C.1 KIIS Survey Data and Variable Definitions 21

C.2 Google Trends Data: Language Use and Cuisine Interest 22

C.3 Baby Names: Data Construction and Validation 25

C.4 Music Consumption: Spotify Charts and Artist-Level Data 31

C.5 Administrative Reform and Geographic Units 33

C.6 Construction of Measures of Internally Displaced Persons 33

D Additional Results **35**

A Literature Review

A.1 Theories of War, Nation-Building, and Identity

The relationship between nation-building and military conflict has received sustained attention across 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 conflict raises demands for taxation and conscription, prompting states to expand administrative capacity, provide public goods, and promote a common identity capable of sustaining mass mobilization (Tilly, 1975; Mann, 1993). Related work emphasizes how industrialization and modernization create pressures for linguistic and cultural standardization, fostering common languages and "high cultures" that facilitate national integration (Gellner, 1983; Anderson, 1983). From these perspectives, nationalism is largely a product of modernity, shaped by the demands of governance and war (Smith, 1998; Hutchinson, 2017).

Building on this tradition, formal economic models treat nation-building as a strategic policy choice. Alesina et al. (2020) argue that governments facing external conflict expand public goods and invest in a common identity to sustain mobilization. Consistent with this mechanism, Aghion et al. (2019) document that external military rivalry increases investment in mass primary education, particularly in democracies. Alesina et al. (2021) instead emphasize domestic political incentives: rulers have stronger incentives to homogenize the population when democratization threatens elite control, again using mass education as a central instrument. More generally, Almagro and Andrés-Cerezo (2020) show theoretically that the success of nation-building in displacing regional identities depends on the surrounding economic and political conditions.

Ethnosymbolist approaches offer a complementary, bottom-up account. Rather than viewing nations primarily as modern constructions, they emphasize the persistence of preexisting myths, symbols, and collective memories. Smith (1986) argues that these inherited materials provide the foundations for durable national identities, while shared trauma and commemoration can strengthen them through collective experience rather than elite instrumentalization (Smith, 1981; Hutchinson, 2017). War may therefore reshape

identity not only through state-led investments but also through the meanings individuals attach to shared wartime experiences.

A related theoretical literature makes identification an individual-level choice. In Shayo (2009), identification depends positively on a group's relative status and negatively on the individual's perceived distance from its prototype, with identity and behavior jointly determined. Sambanis and Shayo (2013) embed this mechanism in a model of ethnic conflict, showing how conflict and identification can reinforce one another and produce either common identification or polarization. Relatedly, Sambanis et al. (2015) emphasize military outcomes as a source of national status, arguing that victory can strengthen national identification while defeat can weaken it. Our conceptual framework draws on this class of models (reviewed in Shayo, 2020; Gidron and Shayo, 2026).

This utility structure is also closely related to Akerlof and Kranton (2000), who model identity as depending on the status of one's social category, congruence between one's attributes and the category ideal, and conformity to its behavioral prescriptions. Kranton (2016) distinguishes short-run behavioral adjustment under given categories and norms from longer-run processes through which those categories and norms evolve. Bénabou and Tirole (2011) provide a distinct mechanism for behavioral feedback: individuals infer their own values partly from past choices, so identity-consistent actions can strengthen the corresponding self-conception through self-signaling. Finally, Akerlof (2016) explains how a common identity can facilitate collective action by inducing individuals to adopt the group's objectives as their own.

Together, these perspectives motivate the central elements of our framework. War can reshape identification by changing group status, the salience of competing boundaries, and individuals' perceived distance from national prototypes. Identity-linked behavior may both express identification and reinforce it by reducing that distance, while stronger common identification may support collective resistance. Because war can move these forces in different directions, its effect on national identity is not determined *ex ante*: it may produce consolidation, polarization, or movement toward a competing identity.

A.2 Policies and Institutions of National Integration

A broad empirical literature examines how policies and institutions shape the balance between national and narrower ethnic or regional identities. Education and language policy are prominent instruments. 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 regional identity, even among ethnic Spaniards. By contrast, exploiting discontinuities in state-sponsored education under France's July Monarchy, Blanc and Kubo (2025) show that public schooling and a state-designed curriculum promoted linguistic homogenization, increased the salience of national identity, and strengthened preferences for political centralization.

A complementary approach emphasizes intergroup contact rather than direct assimilation policies. 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. Other studies examine shared experiences and the economic foundations of integration. Depetris-Chauvin et al. (2020) demonstrate that football victories in sub-Saharan Africa raise national identification and reduce interethnic violence. Beyond explicit nation-building policies, 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.

A related 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 similar spirit, Bazzi et al. (2026) study Indonesia's mass public-school construction program and show that Islamic schools expanded in response rather than being displaced. This competitive response weakened the nation-building effects of public schooling and produced lasting political and cultural consequences, including greater religiosity.

A.3 Conflict, Threat, and Identity

Most directly related to our study is work on how conflict and foreign intervention reshape political allegiance and identity. Dell and Querubin (2018) show that U.S. bombing in Vietnam increased communist insurgent activity, weakened local governance, and reduced noncommunist civic engagement; a firepower-intensive counterinsurgency strategy also worsened attitudes toward the United States and South Vietnamese government relative to a "hearts and minds" approach. While their findings show how external violence can alter political allegiance, they do not study changes in national identity per se. Our paper instead asks whether external war changes national identification and produces identity consolidation or polarization within the attacked society.

DellaVigna et al. (2014) find that exposure to cross-border nationalist Serbian radio in Croatia increased support for extreme nationalist parties and ethnically offensive expression, showing how nationalist media in a postconflict setting can intensify out-group hostility. Relatedly, Cannella et al. (2024) examine the de facto Nazi annexation of parts of northern Italy and find persistent increases in support for radical opposition parties at both ends of the political spectrum. Evidence from the Troubles likewise shows how state violence can entrench the identity of the targeted group: Hughes (2011) argues that British repression reinforced Irish nationalism and contributed to political radicalization.

A complementary literature clarifies the individual-level mechanisms through which threat may affect identity. Brewer (1999) distinguishes attachment to the in-group from hostility toward out-groups, emphasizing that the former need not imply the latter, while de Figueiredo Jr and Elkins (2003) show that different forms of national attachment are differently associated with prejudice. Building on this distinction, Fritsche et al. (2017) argue that threatening events can strengthen in-group identification by restoring a sense of control. Weisel and Zultan (2016) similarly show experimentally that individuals who perceive their group as the target of threat reallocate effort and loyalty toward the in-group. Sambanis and Lee (2023) find that external threat can raise national identification primarily through greater hostility toward the out-group rather than stronger attachment to the national in-group. Complementary evidence from wartime settings indicates that exposure to violence often increases local cooperation and civic engagement, although

these responses tend to be parochial rather than universal (Bauer et al., 2016).

The identity consequences of conflict, however, are not uniformly nationalizing. Dehdari and Gehring (2022) find that exposure to war, foreign occupation, and repression in Alsace-Lorraine strengthened regional identity and demands for regional autonomy. External threat may also redirect identification across political levels: Gehring (2021) shows that Russia's 2014 aggression against Ukraine strengthened European identification in more-threatened parts of Europe while leaving lower-level identities unchanged. Related evidence points to fragmentation through a decline in national status: Ananyev and Poyker (2023) find that proximity to the Tuareg rebellion in Mali reduced national identification and strengthened ethnic identity even among groups not involved in the rebellion, consistent with conflict weakening identification by revealing low state capacity. Qian and Tabellini (2026) provide complementary evidence that political exclusion can weaken national cohesion under external threat. After Pearl Harbor, voluntary enlistment rose less among Black men, especially in counties with greater racial discrimination, consistent with exclusion weakening identification with and support for the state.

Taken together, this literature shows that conflict can strengthen national identification, entrench subnational attachment, or redirect identification toward a broader political community. The direction of identity change therefore cannot be determined *ex ante*, motivating our analysis of when external war produces national consolidation or polarization.

A.4 Measuring National Identity

Recent work complements traditional surveys with behavioral measures of national identity. Social media provides high-frequency evidence, with changes in language use and online self-presentation serving as indicators of identity (Abramenko et al., 2024). Similarly, Metzger et al. (2016) analyze Twitter language choice during the Euromaidan protests, using preferred and actual posting languages as proxies for ethnolinguistic identification.

Naming choices provide a complementary, durable measure. Because names are publicly observable and costly to reverse, changes in their ethnic distinctiveness can reveal shifts in national identity (Fouka, 2019; Esposito et al., 2023; Assouad, 2025).[1]

[1]Longitudinal records can also reveal changes in identity classification. Using linked U.S. census records from 1880 to 1940, Dahis et al. (2019) document that many men recorded as Black in one census were recorded

Consumption behavior offers a third approach. 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 greater salience of national identity shifts demand toward domestic goods. Related work finds that conflict-related shocks during the Iraq War altered U.S. consumers' demand for foreign food brands (Pandya and Venkatesan, 2016; Helms et al., 2025). Closest to our music analysis, Kikuchi (2024) uses classical concert programs from 1900–1960 to show that the two world wars reduced performances of music from enemy countries, especially modern works more closely associated with those nations. In contrast, we show that the reallocation toward Ukrainian-language music extends to songs by Ukrainian artists, distinguishing this response from a boycott of enemy-country artists.

Our paper combines and extends these approaches within a single setting, using high-frequency language choices, private cultural behavior, and durable naming decisions to triangulate a common pattern of national consolidation during wartime.

as white in the next, with validation exercises consistent with racial passing rather than classification error.

B Conceptual Framework: Details

This appendix derives the theoretical implications of the assumptions in Section 1. We consider the two-group case $j \in \{N, R\}$. Following Shayo (2009), we distinguish group *membership*, an objective attribute, from *identification*, the endogenous weight an individual places on a group. We focus on individuals for whom both the Ukrainian and Russian identities are potentially relevant, and who may therefore place positive weight on either or both. This does not preclude a corner solution, in which an individual identifies wholly with one group, and we analyze that possibility as well. Our focus here is on the individual response to the conflict-related shocks. The full Social Identity Equilibrium à la Sambanis and Shayo (2013), endogenizing the intensity and shape of conflict itself (and hence status and salience) is beyond the scope of the empirical analysis in the current paper.

B.1 Setup

From Section 1, the utility that individual i derives from identifying with group j is:

$$I_{ij} = \gamma\, S_j - (1-\gamma)\, d_{ij}, \qquad \gamma \in (0,1) \tag{A1}$$

Let α_{iN} and α_{iR} denote the weights individual i places on the Ukrainian and Russian identities, respectively, with $\alpha_{iN} + \alpha_{iR} = 1$. To save on notation, we write $\alpha_i \equiv \alpha_{iN}$, so that $\alpha_{iR} = 1 - \alpha_i$. The individual chooses these identity weights together with a vector of behavioral adjustments that reposition her in attribute space. Attribute k enters perceived distance through $\delta_k(i,j) \geq 0$, the gap between i and group j's prototype on dimension k. On each dimension, let b_{ik} denote the size of behavioral adjustment, where a positive b_{ik} is a movement toward the national prototype and away from the Russian one.[2] After adjustments, perceived distances are thus

$$d_{iN} = \sum_{k=1}^{K} w_k \underbrace{\left(\delta_k^0(i,N) - b_{ik}\right)}_{\equiv \delta_k(i,N)} \qquad d_{iR} = \sum_{k=1}^{K} w_k \underbrace{\left(\delta_k^0(i,R) + b_{ik}\right)}_{\equiv \delta_k(i,R)} \tag{A2}$$

where $\delta_k^0(i,j)$ is the baseline (preadjustment) gap, $\delta_k(i,j)$ is the effective (postadjustment)

[2] We focus on identity-differentiating adjustments. See Dekel (2026) for an analysis of movements common to both prototypes, i.e., a change that lowers (or increases) both distances at once.

gap, $w_k \geq 0$ is the salience (attention weight) of dimension k, and $\sum_k w_k = 1$.[3]

Adjusting attribute k is costly: let $c_{ik}(b_{ik})$ denote the cost to individual i of an adjustment of size b_{ik} in attribute k. We assume c_{ik} is strictly convex and minimized at $b_{ik} = 0$, and hence is increasing for $b_{ik} > 0$ and decreasing for $b_{ik} < 0$. Marginal costs can vary across dimensions (changing everyday language requires more effort than stating a different ethnicity), as well as across individuals on the same dimension (language is easier to change when young). Using Equation (3) the individual then solves

$$\max_{\alpha_i \in [0,1],\, \{b_{ik}\}} U_i = \pi_i^0 + g(\alpha_i)\, I_{iN} + g(1-\alpha_i)\, I_{iR} - \sum_k c_{ik}(b_{ik}) \tag{A3}$$

where π_i^0 is material payoff before adjustment costs (so that $\pi_i = \pi_i^0 - \sum_k c_{ik}(b_{ik})$ in (3)), $g(0) = 0$, $g' > 0$, and for the remainder of this appendix we assume $g'' < 0$.[4]

B.2 Identity Choice

Consider first the choice of α_i, holding adjustments fixed. When both identification payoffs I_{iN} and I_{iR} are positive, an interior optimum equates the marginal identification returns across the two groups:

$$\frac{g'(1-\alpha_i^*)}{g'(\alpha_i^*)} = \frac{I_{iN}}{I_{iR}} \tag{A4}$$

From concavity of g, we have $\partial \alpha_i^* / \partial I_{iN} > 0$ and $\partial \alpha_i^* / \partial I_{iR} < 0$. In words, the weight of the Ukrainian national identity increases with the identity payoff it provides relative to the Russian identity.

Note that an interior optimum requires that both identity payoffs are positive—if either is nonpositive (or if one is sufficiently large relative to the other), the individual is at a corner. A group's identity payoff can in principle be negative if its status is too low relative to the individual's distance from its prototype. In particular, an individual sufficiently close to the national prototype and far from the Russian one might have $I_{iR} \leq 0 < I_{iN}$, and place all weight on the nation, $\alpha_i^* = 1$. In this case, a further rise in I_{iN} leaves α_i^*

[3] Distance cannot be negative, hence adjustment is bounded $-\delta_k^0(i,R) \leq b_{ik} \leq \delta_k^0(i,N)$.

[4] Strict concavity allows interior identity weights, so comparative statics at an interior optimum are obtained by differentiation. By contrast, when $g(\alpha) = \alpha$, utility is linear in identity weights and the optimum is generically a corner: the individual places all weight on whichever identity yields higher utility. In that case, shocks affect aggregate identification by inducing some individuals to switch between the two corner solutions, so the corresponding comparative statics concern population shares rather than within-individual changes in identity weights (see Atkin et al., 2021).

unchanged. Similarly, $I_{iN} \leq 0 < I_{iR}$ implies $\alpha_i^* = 0$.

B.3 Behavioral Adjustment

Now consider the adjustment margins $\{b_{ik}\}$. Differentiating (A3) with respect to b_{ik}, using (A1) and (A2), the first-order condition for an interior b_{ik} is

$$w_k (1-\gamma) \left[g(\alpha_i) - g(1-\alpha_i) \right] \;=\; c'_{ik}(b_{ik}) \tag{A5}$$

The left-hand side is the return to moving attribute k toward the national prototype. Moving one unit lowers d_{iN}, thereby raising I_{iN}, while raising d_{iR}, thereby lowering I_{iR}. The gain in I_{iN} enters utility with weight $g(\alpha_i)$, whereas the loss in I_{iR} enters with weight $g(1-\alpha_i)$. Each effect is scaled by salience w_k (a movement on a dimension one pays little attention to has little identity payoff) and the distance weight $1-\gamma$. The return to adjustment is therefore proportional to $[g(\alpha_i) - g(1-\alpha_i)]$, which is increasing in α_i, and positive if and only if $\alpha_i > \frac{1}{2}$. That is, adjustment toward the national prototype is worthwhile only for individuals who place more weight on the Ukrainian than on the Russian-leaning identity.[5] An interior solution $b_{ik} > 0$ is optimal whenever the return to the first unit of adjustment exceeds its marginal cost:

$$b_{ik} > 0 \iff w_k (1-\gamma) \left[g(\alpha_i) - g(1-\alpha_i) \right] \;>\; c'_{ik}(0^+) \tag{A6}$$

where $c'_{ik}(0^+)$ is the marginal cost of the first positive unit of adjustment. If c_{ik} is smooth at zero, so that $c'_{ik}(0) = 0$, then (A6) reduces to $b_{ik} > 0 \iff \alpha_i > \frac{1}{2}$. The individual adjusts toward the national prototype if she identifies more strongly with Ukrainian identity, and toward the Russian prototype if she identifies more strongly with the Russian-leaning identity. However, if there is a cost to adjusting behavior, such that the cost function has a kink at zero, i.e., $c'_{ik}(0^+) > 0$, then small returns leave $b_{ik} = 0$. In this case, behavioral adjustment begins only once the return exceeds the marginal cost of the first unit. Costly behaviors are then adjusted only when the identity shift is large enough. For dimensions that do adjust ($b_{ik} \neq 0$), convexity of c_{ik} implies that the size of the adjustment $|b_{ik}|$ increases with the strength of the dominant identity.

[5] For an individual who places more weight on the Russian-leaning identity ($\alpha_i < \frac{1}{2}$), the return is negative and $b_{ik} < 0$, i.e., they adjust toward the Russian prototype and away from the national one.

B.4 Comparative Statics

Using (A4) and (A5), we can now trace the effects of shocks to each of the three forces: status, distance, and salience. Shocks to status and to distance in attribute space do not enter (A5) directly, so they affect behavioral adjustment only through the identity weight. Their effects are hence relatively straightforward. Salience shocks are slightly more subtle. In all three cases, however, identity and behavior are chosen jointly: a shock that moves the identity weight also moves behavior, and the change in behavior feeds back on the identity weight. We therefore derive the comparative statics from the full system of first-order conditions.

Notation: since the analysis in this subsection concerns a single individual, we suppress the subscript i for this subsection (e.g., write U for U_i, α for α_i, and $\delta_k(j)$ for $\delta_k(i,j)$).

The system of first-order conditions. At an interior optimum, $(\alpha^*, b_1^*, \ldots, b_K^*)$ satisfies

$$\begin{aligned} U_\alpha &\equiv g'(\alpha)\, I_N - g'(1-\alpha)\, I_R = 0, \\ U_{b_k} &\equiv (1-\gamma)\, w_k \big[g(\alpha) - g(1-\alpha)\big] - c_k'(b_k) = 0, \qquad k = 1, \ldots, K \end{aligned} \tag{A7}$$

The first condition is (A4) rearranged, and the second is (A5). Together they implicitly define α^* and $b_1^*, \ldots, b_K^*$ as functions of the parameters: the statuses S_N and S_R, the baseline gaps $\{\delta_k^0(j)\}$, and the attention weights $\{w_k\}$. Let θ denote any one of these parameters. Totally differentiating (A7) with respect to θ then yields

$$\begin{aligned} U_{\alpha\alpha}\frac{d\alpha^*}{d\theta} + \sum_{l=1}^{K} U_{\alpha b_l}\frac{db_l^*}{d\theta} &= -U_{\alpha\theta}, \\ U_{b_k\alpha}\frac{d\alpha^*}{d\theta} + \sum_{l=1}^{K} U_{b_k b_l}\frac{db_l^*}{d\theta} &= -U_{b_k\theta}, \qquad k = 1, \ldots, K \end{aligned} \tag{A8}$$

where subscripts denote partial derivatives of U, evaluated at the optimum. From (A7), U_{b_k} depends on α and on b_k, but on no other b_l: identity payoffs are linear in the adjustments, and costs are separable across dimensions. Hence $U_{b_k b_l} = 0$ for $l \neq k$ and $U_{b_k b_k} = -c_k''$. Using $U_{b_k\alpha} = U_{\alpha b_k}$, the system simplifies to

$$\begin{aligned} U_{\alpha\alpha}\frac{d\alpha^*}{d\theta} + \sum_{l=1}^{K} U_{\alpha b_l}\frac{db_l^*}{d\theta} &= -U_{\alpha\theta}, \\ U_{\alpha b_k}\frac{d\alpha^*}{d\theta} - c_k''\frac{db_k^*}{d\theta} &= -U_{b_k\theta}, \qquad k = 1, \ldots, K \end{aligned} \tag{A9}$$

Solution. From the second row in (A9) we have

$$\frac{db_k^*}{d\theta} = \frac{1}{c_k''}\left[U_{\alpha b_k}\frac{d\alpha^*}{d\theta} + U_{b_k\theta}\right] \tag{A10}$$

Substituting into the first equation of (A9) we can solve for $d\alpha^*/d\theta$:

$$\frac{d\alpha^*}{d\theta} = \frac{1}{D}\left[U_{\alpha\theta} + \sum_l \frac{U_{\alpha b_l}}{c_l''}\,U_{b_l\theta}\right], \qquad D \equiv -U_{\alpha\alpha} - \sum_l \frac{U_{\alpha b_l}^2}{c_l''} \tag{A11}$$

At a regular interior maximum we have $D > 0$.[6] Hence, the sign of this derivative is the sign of the expression in brackets. The first term in brackets is the direct effect of θ on the marginal return to identification. The second term is the effect of θ via its effect on the behavioral adjustments $(b_1^*, \ldots, b_K^*)$. θ shifts behavior directly (by changing the net return to behavioral adjustments through $U_{b_l\theta}$), and the change in behavior shifts the marginal return to identification (through $U_{\alpha b_l}$).

Turning to the denominator D, differentiating U_α in (A7) we have

$$U_{\alpha\alpha} = g''(\alpha^*)\,I_N + g''(1-\alpha^*)\,I_R, \qquad U_{\alpha b_k} = (1-\gamma)\,w_k\left[g'(\alpha^*) + g'(1-\alpha^*)\right] \tag{A12}$$

Since $g'' < 0$, and at an interior optimum both I_N and I_R are positive, $U_{\alpha\alpha} < 0$. Next, on any dimension that receives attention ($w_k > 0$), $U_{\alpha b_k} > 0$. This is the complementarity between identification and behavior: moving attribute k toward the national prototype raises I_N and lowers I_R, and hence raises the marginal return to national identification. Finally, note that $D < -U_{\alpha\alpha}$, since each term $U_{\alpha b_l}^2/c_l''$ is positive.

Shocks to status and to distance in attribute space. Status and baseline distances do not enter U_{b_k} in (A7), so $U_{b_k\theta} = 0$ for these parameters, and (A11) and (A10) reduce to

$$\frac{d\alpha^*}{d\theta} = \frac{U_{\alpha\theta}}{D}, \qquad \frac{db_k^*}{d\theta} = \frac{U_{\alpha b_k}}{c_k''}\,\frac{d\alpha^*}{d\theta} \tag{A13}$$

[6] We consider a regular interior maximum, at which the Hessian of (A3) with respect to $(\alpha, b_1, \ldots, b_K)$ is negative definite; this is also what the implicit function theorem requires. That Hessian is the coefficient matrix of (A9), and its determinant equals $(-1)^{K+1} D \prod_k c_k''$, so given $c_k'' > 0$ negative definiteness is equivalent to $D > 0$. Note that $g'' < 0$ and convex costs do not by themselves deliver it: if adjustment is cheap enough, the complementarity can be strong enough that $\sum_l U_{\alpha b_l}^2/c_l'' \geq -U_{\alpha\alpha}$. The sums over l run over dimensions at which (A5) holds with equality. A dimension at a corner—held at $b_l^* = 0$ by a kink in the cost function, as in (A6), or at a bound of the adjustment interval, where the effective gap from one of the prototypes reaches zero—drops out of (A9). Its adjustment does not respond to small status or salience shocks, which leave the corner in place. A shock to the baseline gap that defines a bound is different, since it moves the bound itself.

Notice that if we held behavior fixed, the response of the identity weight to the shock would be $U_{\alpha\theta}/(-U_{\alpha\alpha})$. Since $0 < D < -U_{\alpha\alpha}$, the *total* response has the same sign but is larger in absolute value. This reflects a positive feedback: a higher α^* raises the return to adjustment; adjustment lowers the distance from the national prototype and raises the distance from the Russian one; and this further raises α^*.

Consider status first. An increase in national status S_N raises I_N. Since $U_{\alpha S_N} = \gamma\, g'(\alpha^*) > 0$, this raises α^*. Similarly, α^* rises if Russian status falls, since $U_{\alpha S_R} = -\gamma\, g'(1-\alpha^*) < 0$. From (A13), every adjustment b_k^* on a dimension that receives positive attention moves in the same direction as α^*. The increase is strict on dimensions at an interior optimum; on a dimension where adjustment is held at zero by a kink in the cost function, adjustment begins only once the return exceeds the threshold in (A6).

Turning to distance in attribute space, consider an exogenous shock to some baseline gap $\delta_k^0(j)$. A fall in $\delta_k^0(N)$, the baseline distance from the national prototype on dimension k, raises I_N and hence the marginal return to national identification: $U_{\alpha\delta_k^0(N)} = -(1-\gamma)\, w_k\, g'(\alpha^*) < 0$. Similarly, an increase in the baseline distance from the Russian prototype lowers I_R: $U_{\alpha\delta_k^0(R)} = (1-\gamma)\, w_k\, g'(1-\alpha^*) > 0$. Both shocks thus increase α^*, provided adjustment on the shocked dimension k is interior.[7] Furthermore, by (A13) every adjustment b_l^* on a dimension that receives attention weakly increases. Thus, a closing of the gap from the national prototype on one dimension leads to further endogenous adjustments on other dimensions. And on dimension k itself, the effective gap $\delta_k(N) = \delta_k^0(N) - b_k^*$ falls by more than the initial shock.

Shocks to salience. Changing attention weights has two effects. First, a direct effect: a higher weight on dimension k scales up the return to behavioral adjustment on that dimension, as w_k enters U_{b_k} directly. Second, an identity effect: paying more attention to some dimensions and less to others changes perceived distances. This affects the identity payoffs I_N and I_R and hence α^*. Whether this raises national identification depends on

[7] A shock to $\delta_k^0(j)$ also moves the adjustment bounds of footnote 3. Thus, if the individual has completely closed the gap to the national prototype ($b_k^* = \delta_k^0(N)$ so that $\delta_k(N)$ is zero), then a fall in $\delta_k^0(N)$ actually lowers b_k^*. This means the effective national gap stays at zero while the effective Russian gap $\delta_k^0(R) + b_k^*$ falls, raising I_R and lowering α^*. Note that this possibility arises only for shocks to baseline distance: status and salience shocks leave the bounds unchanged, so the comparative statics we take to the data are unaffected. The spillover to other dimensions is also unaffected, since their bounds do not move.

where the newly salient dimensions place the individual relative to the two prototypes.

Since attention is finite ($\sum_k w_k = 1$), increasing the weight on one dimension must reduce the weight on others. In other words, a salience shock is a reallocation of attention across dimensions. We can write it as

$$w_k(q) = w_k + q\, m_k, \qquad \sum_k m_k = 0, \tag{A14}$$

where q is the size of the shock and $m = (m_1, \ldots, m_K)$ is its profile: dimensions with $m_k > 0$ gain attention, and those with $m_k < 0$ lose it. Note that in general, several dimensions can gain attention (for instance, both citizenship and country of residence), while others can lose different amounts (and some may have $m_k = 0$).[8]

We can now examine the effect of the salience shock by differentiating with respect to q at $q = 0$, that is, at the preshock attention weights. Holding behavior fixed, the reallocation of attention changes *perceived* distances by

$$\frac{\partial d_N}{\partial q} = \sum_k m_k\, \delta_k(N), \qquad \frac{\partial d_R}{\partial q} = \sum_k m_k\, \delta_k(R), \tag{A15}$$

where $\delta_k(j)$ are the effective gaps in (A2). Because $\sum_k m_k = 0$, $\partial d_N / \partial q$ is proportional to the average gap from the national prototype on the dimensions gaining attention minus the average gap on the dimensions losing attention (each average weighted by $|m_k|$). It is negative when the reallocation shifts attention toward dimensions on which the individual is relatively close to the national prototype, so she perceives herself as more similar to it.

From (A7), the two cross-partials are $U_{\alpha q} = -(1-\gamma)\left[g'(\alpha^*)\, \partial d_N/\partial q - g'(1-\alpha^*)\, \partial d_R/\partial q\right]$ and $U_{b_k q} = (1-\gamma)\, m_k \left[g(\alpha^*) - g(1-\alpha^*)\right]$. Notice that unlike status and distance, salience enters the behavioral condition directly, so $U_{b_k q} \neq 0$. Substituting into (A11) and us-

[8] Since weights are nonnegative, $m_k \geq 0$ on any dimension with $w_k = 0$. The scale of m is arbitrary: rescaling m and q inversely describes the same shock.

ing (A12),

$$\frac{d\alpha^*}{dq} = \frac{1-\gamma}{D}\Bigg\{ \underbrace{-g'(\alpha^*)\frac{\partial d_N}{\partial q} + g'(1-\alpha^*)\frac{\partial d_R}{\partial q}}_{\text{identity channel}} + \underbrace{(1-\gamma)\big[g'(\alpha^*)+g'(1-\alpha^*)\big]\big[g(\alpha^*)-g(1-\alpha^*)\big]\sum_l \frac{m_l\, w_l}{c_l''}}_{\text{behavioral feedback}} \Bigg\}. \tag{A16}$$

Since $D > 0$, $d\alpha^*/dq$ has the sign of the term in braces.

The identity channel is the effect of the reallocation on identity payoffs at fixed behavior. It is positive when the reallocation brings the individual perceptually closer to the national prototype ($\partial d_N/\partial q < 0$) and farther from the Russian one ($\partial d_R/\partial q > 0$), with each change weighted by the marginal identification return g'.

The behavioral feedback arises because the reallocation also shifts behavior directly, and the change in behavior shifts identity payoffs. Its sign depends on several factors. The factor $g(\alpha^*) - g(1-\alpha^*)$ is positive if and only if $\alpha^* > \frac{1}{2}$. Because $\sum_l m_l = 0$, the factor $\sum_l m_l w_l/c_l''$ compares the behavioral responsiveness w_l/c_l'' of the dimensions gaining attention with that of the dimensions losing it. For an individual who identifies more with the nation, the feedback is positive when attention moves toward dimensions that are more readily adjusted, and negative when it moves toward dimensions that are not; the reverse holds for an individual who identifies more with the Russian-leaning group. The feedback vanishes as $\alpha^* \to \frac{1}{2}$, and on dimensions held at zero by a kink in the cost function. When the feedback and the identity channel have the same sign, they reinforce each other. When they have opposite signs, the identity response has the sign of the identity channel as long as the latter is larger in absolute value.

In our setting, a Russian attack on the country as a whole increases the salience of dimensions aligned with the interstate cleavage, such as citizenship and country of residence, on which residents are close to the national prototype and far from the Russian one. It lowers the relative weight on attributes that separate ethnic Russians from ethnic Ukrainians, most directly ascribed ethnicity, i.e., ethnic origin by parental descent. For such a reallocation, $\partial d_N/\partial q < 0$ and $\partial d_R/\partial q > 0$, so the identity channel is positive. The behavioral feedback is probably second order here. Attention moves largely between at-

tributes that are very hard to change in the short run: citizenship and country of residence on one side, ascribed ethnicity on the other, which in Ukraine is inherited through parental origin (see Section 2). Thus, c_l'' is large on the dimensions with $m_l \neq 0$, and $\sum_l m_l w_l / c_l''$ is close to zero. The sign of the identity response is then governed by the identity channel. Note that this reallocation does not require the salience of language to fall: (A14) leaves the language dimension free to gain attention, lose attention, or hold the same weight.

Behavioral response to salience shocks. Substituting (A16) into (A10) gives

$$\frac{db_k^*}{dq} = \frac{1-\gamma}{c_k''}\Bigg[\underbrace{m_k\big(g(\alpha^*) - g(1-\alpha^*)\big)}_{\text{direct channel}} + \underbrace{w_k\big(g'(\alpha^*) + g'(1-\alpha^*)\big)\frac{d\alpha^*}{dq}}_{\text{identity channel}}\Bigg]. \tag{A17}$$

Since $(1-\gamma)/c_k'' > 0$, the sign of db_k^*/dq is the sign of the term in brackets. The direct channel has the sign of $m_k(\alpha^* - \frac{1}{2})$: for people identifying more with the nation ($\alpha^* > \frac{1}{2}$), more attention to dimension k increases the return to adjusting attribute k toward the national prototype, and less attention reduces it. The identity channel has the sign of $d\alpha^*/dq$ from (A16).

Table B.1 collects the possible configurations. Since the direct channel has the sign of $m_k(\alpha^* - \frac{1}{2})$ and the identity channel the sign of $d\alpha^*/dq$, the sign of db_k^*/dq is unambiguous when the two agree, and depends on their relative size when they do not.

In the cells where the two channels oppose each other, which one dominates is a quantitative matter: the direct term is proportional to $m_k\big[g(\alpha^*) - g(1-\alpha^*)\big]$ and the identity term to $w_k\big[g'(\alpha^*) + g'(1-\alpha^*)\big]d\alpha^*/dq$, and either can be larger. However, as $\alpha^* \to \frac{1}{2}$ the direct term vanishes, so for individuals close to indifference the identity channel dominates, provided the identity response itself does not vanish.

Considering the implications of the upper panel for our setting, the dimensions whose attention weights change most—citizenship and country of residence upward, ascribed ethnicity downward—are arguably ones that cannot be adjusted much, so they account for little of the behavioral response. The adjustable margins, language and cultural consumption, are those whose weight the reallocation need not change. If so, they fall in the middle row, where only the identity channel operates and adjustment moves with α^* for both groups. Should an adjustable margin lose salience as well, the two channels oppose

Table B.1: Behavioral Response to a Salience Shock

	Individual identifies more with	
	the nation ($\alpha^* > \frac{1}{2}$)	the Russian-leaning group ($\alpha^* < \frac{1}{2}$)
Shock raises α^*		
k gains attention ($m_k > 0$)	$db_k^*/dq > 0$	identity $+$, direct $-$
k unchanged ($m_k = 0$)	$db_k^*/dq > 0$	$db_k^*/dq > 0$
k loses attention ($m_k < 0$)	identity $+$, direct $-$	$db_k^*/dq > 0$
Shock lowers α^*		
k gains attention ($m_k > 0$)	identity $-$, direct $+$	$db_k^*/dq < 0$
k unchanged ($m_k = 0$)	$db_k^*/dq < 0$	$db_k^*/dq < 0$
k loses attention ($m_k < 0$)	$db_k^*/dq < 0$	identity $-$, direct $+$

Notes: The table applies to a dimension k that receives positive attention ($w_k > 0$) and whose adjustment is interior. On a dimension held at a corner, by a kink in the cost function or at a bound of the adjustment interval, b_k^* does not respond to a small salience shock.

each other for individuals who identify more with the nation. For those who identify more with the Russian-leaning group they do not: the shock weakens their Russian identity, and less weight on that margin also lowers their return to adjusting toward the Russian prototype, so both channels move b_k^* upward.

Under smooth costs, b_k^* has the sign of $(\alpha^* - \frac{1}{2})$ on every dimension that receives attention. Hence, if the shock is large enough to switch which identity receives more weight (i.e., α^* crosses $\frac{1}{2}$), the direction of behavioral adjustment reverses with it. Once a Russian identifier becomes a national identifier (i.e., α^* crosses $\frac{1}{2}$ from below), she adjusts toward the *national* prototype ($b_k^* > 0$). At the crossing itself, the two identities are equally strong, so there is no benefit to adjusting toward either prototype and $b_k^* = 0$.

For individuals whom the shock carries across $\frac{1}{2}$, behavior and identification need not move together throughout: near the point of indifference, the direct force can still pull adjustment toward the still-dominant prototype even as identification shifts. Whether this produces a detectable nonmonotonicity depends on the forms of the identification and cost functions and on how far the individual starts from $\alpha^* = \frac{1}{2}$.

B.5 Summary

We now restore the subscript i to the notation, since this summary compares individuals.

Status. Status shocks have unambiguous effects on identification. By (A13), the response has the sign of the corresponding cross-partial, so a rise in S_N or a fall in S_R raises α_i^*. The effect is strict for individuals at interior weights; those at a corner do not respond to small shocks. The size of the effect depends on the marginal identification return: $d\alpha_i^*/dS_N = \gamma\, g'(\alpha_i^*)/D$. Thus, for a given D the response is proportional to $g'(\alpha_i^*)$. Since g is concave, $g'(\alpha_i^*)$ is larger the less attached the individual is to the national identity, and the response is therefore largest *away* from the national corner. Allowing behavior to adjust strengthens the response (the total response exceeds the fixed-behavior response by the factor $(-U_{\alpha\alpha})/D > 1$).

Distance in attribute space. A fall in a baseline gap from the national prototype, or a rise in a baseline gap from the Russian prototype, works in the same way: it raises α_i^*, with the same amplification through behavior, and it induces further adjustment on the remaining dimensions. The one qualification is that adjustment on the shocked dimension must be interior. Where the individual has already closed that gap entirely, the sign can reverse.

Salience. Salience shocks are different because they move behavior directly as well as through identification. Equation (A16) separates the two channels. The identity channel is the effect the framework emphasizes: shifting attention toward dimensions on which the individual is close to the national prototype, and away from dimensions on which she is far from it, raises α_i^*. The behavioral feedback has no unconditional sign; as discussed above, it depends on whether the dimensions gaining attention are more or less readily adjusted than those losing it, and on which identity the individual weights more. In the case relevant here, attention arguably moves mainly between attributes that cannot be adjusted much in the short run, so the feedback is probably second order and the sign of the identity response is then governed by the identity channel.

Two features of the salience response matter for what follows.

First, consider two reallocations of attention that are exact opposites, m and $-m$.

Starting from the same point, the marginal identity response to the second is exactly minus the response to the first, and the same holds for each behavioral response.[9] The sign of the identity response is therefore not a property of salience shocks as such. It is a property of the direction in which attention moves, together with where the individual sits on the dimensions whose weight changes. A war that makes the interstate boundary salient and one that makes the *within*-Ukraine ethnolinguistic boundary salient move α_i^* in opposite directions. The framework thus makes no presumption in favor of consolidation. That a war raises the salience of group boundaries is not in itself an argument for national consolidation; what matters is which boundaries it makes salient, and where residents stand on them.

Second, the response varies across individuals. It depends on the marginal identification returns $g'(\alpha_i^*)$ and $g'(1-\alpha_i^*)$, on where each individual stands on the dimensions whose weight changes, through $\partial d_{iN}/\partial q$ and $\partial d_{iR}/\partial q$, and on the curvature D; and individuals at a corner do not respond at all. The aggregate response to a given reallocation therefore depends on the composition of the population.

Behavior. For status and distance shocks, (A13) implies that adjustment on every dimension with positive attention moves in the same direction as α_i^*. A salience shock also enters the return to adjustment directly through w_k. As (A17) shows, both channels affect the size of adjustment toward the individual's favored prototype. The direct channel implies that greater attention to a dimension increases this adjustment, and less attention reduces it. The identity channel increases it when the shock strengthens relative identification. The two channels reinforce each other when they push the same way: a dimension gaining attention when identification strengthens, or losing attention when it weakens. In the mixed cases they work against each other.

Because c_{ik} varies across dimensions and individuals, the same identity shift yields larger behavioral adjustments on cheaper margins and among individuals who adjust at lower cost. Finally, adjustment changes the effective gaps and thereby feeds back into α_i^*.

[9] Both terms in the braces of (A16) are linear in m: the identity channel through $\partial d_{iN}/\partial q = \sum_k m_k \delta_k(i,N)$ and $\partial d_{iR}/\partial q$, and the behavioral feedback through $\sum_l m_l w_l / c_l''$. The denominator D does not involve m. Hence $d\alpha_i^*/dq$ is linear, and so odd, in m: $d\alpha_i^*/dq\big|_{-m} = -\,d\alpha_i^*/dq\big|_m$.

Taking the model to the data. Three points are worth highlighting when taking these comparative statics to the data. First, the identity weight α_i^* and the behavioral adjustments b_{ik}^* are latent variables: neither is observed directly in the data. For α_i^*, we observe a discrete self-report of national identification. For the b_{ik}^*, we observe several behavioral margins, of which language—one of the central margins—is itself recorded as a discrete indicator. We can, however, think of these observed variables as threshold variables. If z_i^* is the relevant latent object (α_i^* or a b_{ik}^*), then the observed outcome is an indicator $y_i = \mathbf{1}\{z_i^* \geq \tau_i\}$, where τ_i is the threshold. The empirical predictions in the main text build on this mapping from the model's latent objects to their observed, largely discrete, counterparts.

Second, the comparative statics apply to individuals for whom both identities are potentially relevant. Individuals whose identification is already strongly anchored in one identity have less scope to respond to changes in relative status or salience. This has an observable implication: districts with higher prewar ethnic-Russian shares are likely to contain more individuals for whom the Russian-leaning identity is relevant, so the responses predicted by the model should be more pronounced there.

Third, a larger identity shock can show up in costly behavioral adjustments that are not observed under a smaller shock. This can arise for two distinct reasons. One is the adjustment friction in (A6): a kink in the cost function at zero. This means that $b_{ik}^* = 0$, as long as the return does not exceed the marginal cost of the first unit. A costly margin thus adjusts only when the shock is large enough. The other reason is measurement. Even where the latent adjustment is positive, a small shock may move it too little to change the observed discrete outcome. Either way, adjustment on costly dimensions is observed only for sufficiently large shocks.

C Additional Details on Data Construction

C.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 Ukraine's adult population. Except for Table D.13, which examines differential patterns in occupied Donbas, we restrict the sample to respondents residing in territories under Ukrainian government control at the time of each wave.

C.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 also 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 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 in which the interview is conducted, following a standardized protocol that leaves the choice to the respondent. Interviewers begin with a greeting whose language is deliberately unmarked, then ask the language question in both languages, leading with the language in which the respondent answered the greeting:

> ***Instructions:*** *Greet the respondent* without accentuating the language of the greeting, *in order to establish in which language the respondent is more comfortable speaking with you. If the respondent replies in Ukrainian, ask: "Tell me, please, is it easier for you to speak Ukrainian,* or perhaps it is easier for you to speak Russian?" *(with the first half in Ukrainian and the second in Russian: "*Скажіть, будь ласка, Вам легше розмовляти українською мовою, или, может быть, Вам легче разговаривать по-русски?*") If the respondent replies in Russian, ask the mirrored question, beginning in Russian and ending in Ukrainian. Take the questionnaire in the language named and conduct the interview in that language. If the respondent expresses no preference, ask which of the two languages he or she speaks more often.*

The Ukrainian and Russian instruments are full parallel translations, the language item appears first in both, and a response is mandatory. Because both languages are explicitly offered through a bilingual prompt, the recorded interview language reflects the respondent's own choice rather than the interviewer's. Interviewers record the language actually used during the interview as Ukrainian, Russian, or mixed Ukrainian–Russian speech (*Surzhyk*). This outcome captures respondents' language choice in a low-stakes interaction, a behavior that is nevertheless costlier to adjust than stated ethnic identification.

C.2 Google Trends Data: Language Use and Cuisine Interest

C.2.1 Language Use in Google Searches

Word-List Construction. To construct comparable Ukrainian–Russian search measures, we begin with Ukrainian- and Russian-language web corpora from the Leipzig Corpora Collection (Goldhahn et al., 2012; Leipzig Corpora Collection, 2025). Each corpus contains 1 million sentences drawn from websites in Ukraine: the Ukrainian-language corpus was compiled in 2019 and the Russian-language corpus in 2015.

We extract nouns from each corpus. 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.[10]

Google Trends Queries and Aggregation. For each word pair, we retrieve oblast-level monthly Google Trends indices for both the Ukrainian and Russian spellings over the January 2012–June 2025 period. Although Google Trends data are inherently relative and scaled within each query, retrieving both spellings simultaneously ensures that the two series are directly comparable. We sum the monthly indices within each quarter and use them to construct the Ukrainian-language search share for each word–oblast–quarter observation. Because the share is defined within each word pair, it is comparable across regions and over time.

Outcome Construction. Let $popularity^{uk}_{iot}$ and $popularity^{ru}_{iot}$ denote the Google Trends search indices for the Ukrainian and Russian spellings of word pair i in oblast o and period t, respectively. We define the Ukrainian spelling share as

$$S_{iot} = 100 \times \frac{popularity^{uk}_{iot}}{popularity^{uk}_{iot} + popularity^{ru}_{iot} + \epsilon}, \quad \epsilon = 10^{-3}$$

Sample. The estimation sample covers 198 word pairs in 21 oblasts over 2012Q1–2025Q2, excluding Crimea, Sevastopol, and the four oblasts Russia claimed to annex in 2022. We drop word–oblast cells with no recorded searches for either spelling throughout the sample. The resulting sample contains 223,938 word–oblast–quarter observations.

Estimation. To construct Panel B of Figure 3, we first residualize S_{iot} using the calendar-quarter–word-pair–oblast fixed effects, removing seasonal and level differences. We then estimate word-pair- and oblast-specific linear trends through 2021Q4, extrapolate them, and subtract the fitted trends. Finally, we regress the adjusted series on quarter indicators. The plotted coefficients measure percentage-point deviations from the seasonally adjusted preinvasion trend. Appendix Figure D.10 instead interacts quarter indicators with the

[10] Examples of word pairs among the ten most common include "year" (Ukrainian: *rik*; Russian: *god*), "person" (Ukrainian: *liudyna*; Russian: *chelovek*), and "child" (Ukrainian: *dytyna*; Russian: *rebenok*).

oblast's prewar share of ethnic Russians over 2019Q1–2025Q2, with word-pair-by-oblast and quarter fixed effects without seasonal or trend adjustment. Both specifications omit 2021Q4. Standard errors are two-way clustered by oblast and word pair.

C.2.2 Interest in National Cuisine

Dish Selection. We measure identity shifts in part through 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 monthly Google Trends indices by oblast from 2012 to 2024. 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. Each Google Trends request jointly retrieves two series: the focal dish and *borscht*, which serves as a common reference term. Because Google Trends places terms included in the same request on a common scale, we scale each dish relative to the borscht series retrieved in the same request. The common borscht anchor makes the resulting indices comparable across requests, allowing us to construct consistent relative popularity distributions across countries (Ukraine and Russia) and over time.

Ukrainianness Score. Each dish i is assigned a time-invariant Ukrainianness score:

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

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

Validation. We assess the external validity of cuisine Ukrainianness score using within-Ukraine cross-sectional variation. Appendix Figure C.1 plots, for each oblast, the average Ukrainianness of dishes implied by prewar search behavior against the oblast's share of ethnic Russians in the 2001 Census. The relationship is strongly negative: oblasts with a larger prewar ethnic Russian presence display systematically lower baseline Ukrainianness

in cuisine-related searches. This pattern is consistent with the intended interpretation of the score and mirrors the validation exercise for baby names.

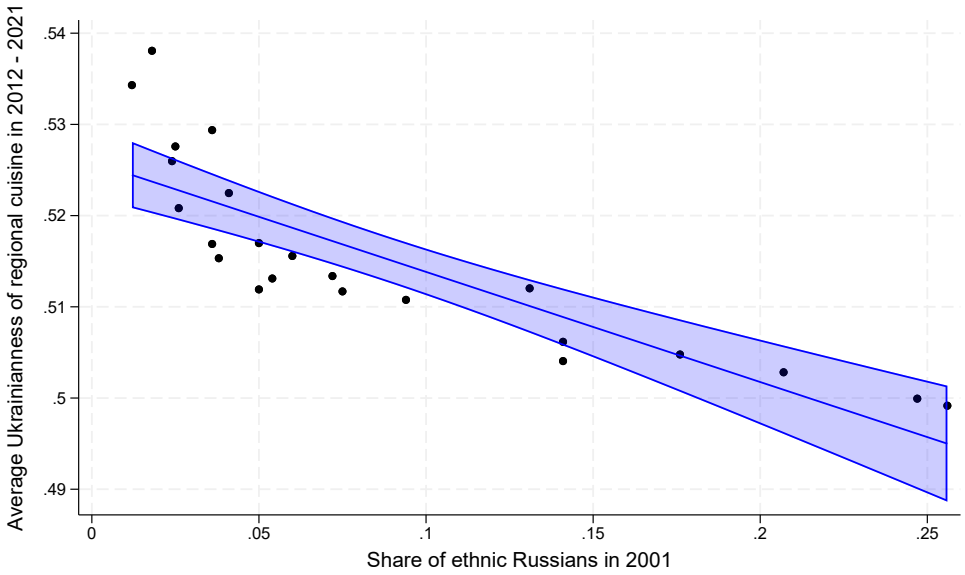


Figure C.1: Validating the Cuisine Ukrainianness Score

Notes: This figure assesses the external validity of the cuisine *Ukrainianness* score within Ukraine. It plots the oblast's average prewar dish Ukrainianness—computed as the search-share-weighted mean of dish-level Ukrainianness using Google Trends data from 2012 to 2021—against the oblast's share of ethnic Russians in 2001 (Census). Oblasts that were at least partially occupied (Crimea, Donetsk, Luhansk, and Sevastopol) are excluded. The solid line is the linear fit; the shaded area is the 95% confidence band. The slope of the fitted line is -0.121 ($p < 0.001$). Consistent with the interpretation of the score, oblasts with a higher prewar share of ethnic Russians tend to display lower Ukrainianness of cuisine-related searches.

Outcomes. Let $\text{Popularity}_{iot}^{raw}$ denote the Google Trends score for dish i in oblast o and month t. We also report a standardized outcome:

$$\text{Popularity}_{iot}^{std} = \frac{\text{Popularity}_{iot}^{raw} - \overline{\text{Popularity}_{io,2012\text{–}2021}^{raw}}}{sd(\text{Popularity}_{io,2012\text{–}2021}^{raw})}$$

computed dish-by-dish using the prewar distribution.

C.3 Baby Names: Data Construction and Validation

Ministry of Justice lists. We collect lists of the most popular baby names by oblast from the Ukrainian Ministry of Justice for 2015 through 2024. The lists are reported annually from 2015 through 2017 and semiannually thereafter. The number of names listed varies across oblast–period cells. The resulting panel has three dimensions: name n, oblast o, and reporting period t.

Name Ukrainianness. Following the approach of constructing ethnic-name indices from relative name frequencies across groups (Fryer and Levitt, 2004), each name is assigned a time-invariant Ukrainianness score based on its prewar relative popularity in Ukraine vs. Russia using *forebears.io*:

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

where U_n and R_n are the prewar shares of name n in Ukraine and Russia, respectively. We harmonize alternative spellings, transliterations, and language-specific variants of the same underlying name under a common identifier (e.g., "Serhii" and "Sergey").

The resulting dataset contains 211 unique names (100 boys' names and 111 girls' names). The Ukrainianness score is approximately uniformly distributed over $[0, 1]$, with first quartile, median, and third quartile equal to 0.21, 0.50, and 0.84, respectively.

Validation. Appendix Figure C.2 plots each oblast's average popular name's Ukrainianness score (computed over the 2015–2021 period) against the oblast's share of ethnic Russians in the 2001 Census. The fitted relationship is negative: regions with a larger prewar 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.

Alternative specification. As a transparent alternative to the main specification, we estimate period effects for the average Ukrainianness of popular names at the oblast–period level:

$$\text{Ukrainianness}_{ot} = \sum_{s \neq 2021\text{ II}} \beta_s \, \mathbf{1}[\text{period} = s] + X_{ot}'\Gamma + \mu_o + \varepsilon_{ot} \tag{A18}$$

where the second half of 2021 is the omitted period and μ_o are oblast fixed effects. Controls X_{ot} include the number of distinct names listed in oblast o and period t (allowing for an oblast-specific slope), as well as oblast-by-period measures of internal displacement inflows and a proxy for refugee outflows. Standard errors are clustered by oblast.

Figure C.3 presents the results: coefficients are near or below zero and stay roughly flat, followed by a sharp and sustained increase beginning in the first half of 2022. This indicates that, adjusting for our controls, the average Ukrainianness of popular names in a

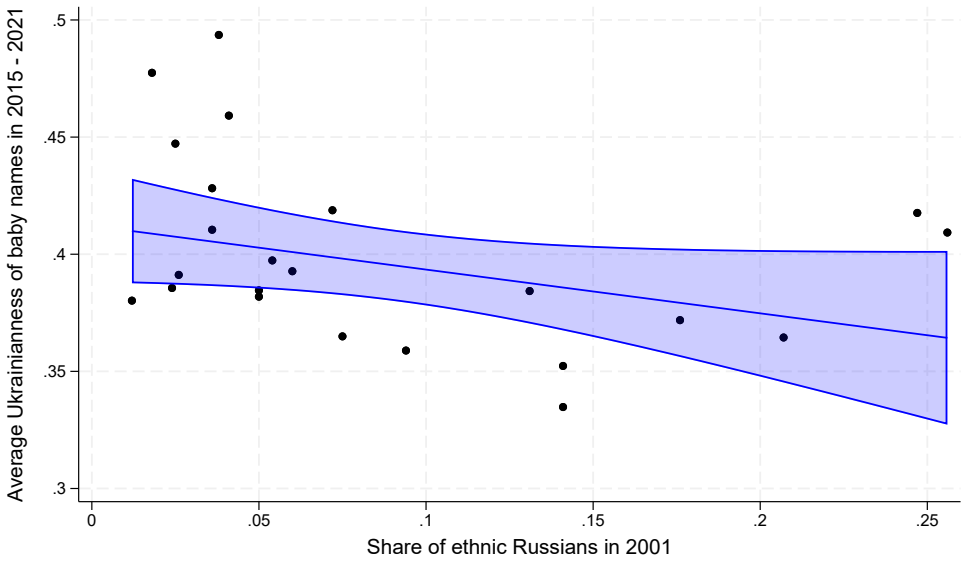


Figure C.2: Validating the Baby-Name Ukrainianness Score

Notes: This figure assesses the external validity of the baby-name *Ukrainianness* score within Ukraine. It plots the oblast's average Ukrainianness of names appearing in the Ministry of Justice popularity lists during the 2015–2021 period against the oblast's share of ethnic Russians in the 2001 Census. Oblasts that were at least partially occupied (Crimea, Donetsk, Luhansk, and Sevastopol) at the time are excluded. The solid line shows the linear fit, weighted by the number of name-list entries per oblast, and the shaded area indicates the 95% confidence band. The slope of the fitted line is -0.187 ($p = 0.072$). Consistent with the interpretation of the score, oblasts with higher prewar shares of ethnic Russians tend to have lower average Ukrainianness of popular baby names.

given oblast rose after the 2022 full-scale invasion.

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

$$\text{Popular}_{nto} = \tau_n \mathbf{1}[t \geq \text{Feb 2022}] + \mu_{to} + \eta_{no} + u_{nto} \tag{A19}$$

where Popular_{nto} is the popularity-list indicator of Equation (6) in the main text, and relate the estimated $\hat{\tau}_n$ to Ukrainianness_n (unweighted and weighted by pre-2022 popularity):

$$\hat{\tau}_n = \alpha + \beta\, \text{Ukrainianness}_n + \varepsilon_n \tag{A20}$$

Figure C.4 shows a positive and statistically significant gradient, indicating that the post-2022 reallocation toward Ukrainian names is not driven by a handful of outliers.

External validation: Ukrainians in Kazakhstan. An alternative interpretation of the shift toward more Ukrainian baby names is that naming choices respond to changing external

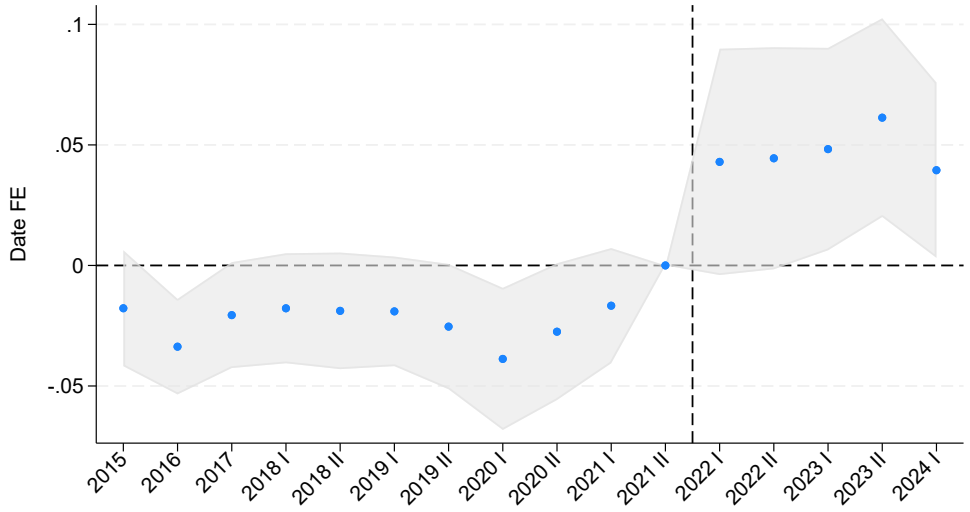


Figure C.3: Average Ukrainianness of Popular Baby Names Over Time

Notes: This figure plots period coefficients from Equation (A18), which estimates changes in the average Ukrainianness of popular baby names across oblasts over time. The outcome, $\text{Ukrainianness}_{ot}$, is the mean Ukrainianness score of names appearing in the Ministry of Justice popular-name list in oblast o and period t. The specification includes oblast fixed effects, oblast-specific slopes in the number of listed names (N_{ot}), and controls for refugee outflows and the share of internally displaced persons (IDPs). Coefficients are normalized to zero in the second half of 2021 and therefore represent deviations relative to the preinvasion period. Standard errors are clustered at the oblast level; whiskers denote 95% confidence intervals.

socioeconomic incentives within Ukraine rather than to changes in intrinsic identity. For example, following the invasion, parents may have become more likely to choose distinctly Ukrainian names if they expect Russian-coded names to carry social or economic costs for their children or otherwise disadvantage them in the future. This possibility makes it informative to examine whether a similar naming shift occurred among Ukrainians living outside Ukraine, where such Ukraine-specific incentives should be substantially weaker.

We examine this possibility using administrative birth records from Kazakhstan (Bureau of National Statistics of Kazakhstan, 2026). These data offer two measurement advantages relative to the Ukrainian Ministry of Justice lists. First, they record the distribution of names among births rather than only whether a name appears on a regional list of popular names. Second, they report children's ethnicity, allowing us to study the evolution of naming choices specifically among ethnically Ukrainian children. Among the 1,608 unique names given to ethnically Ukrainian children over the 2019–2025 period, 189 can be matched to the Ukrainianness scores used in our main analysis. We restrict the analysis to births

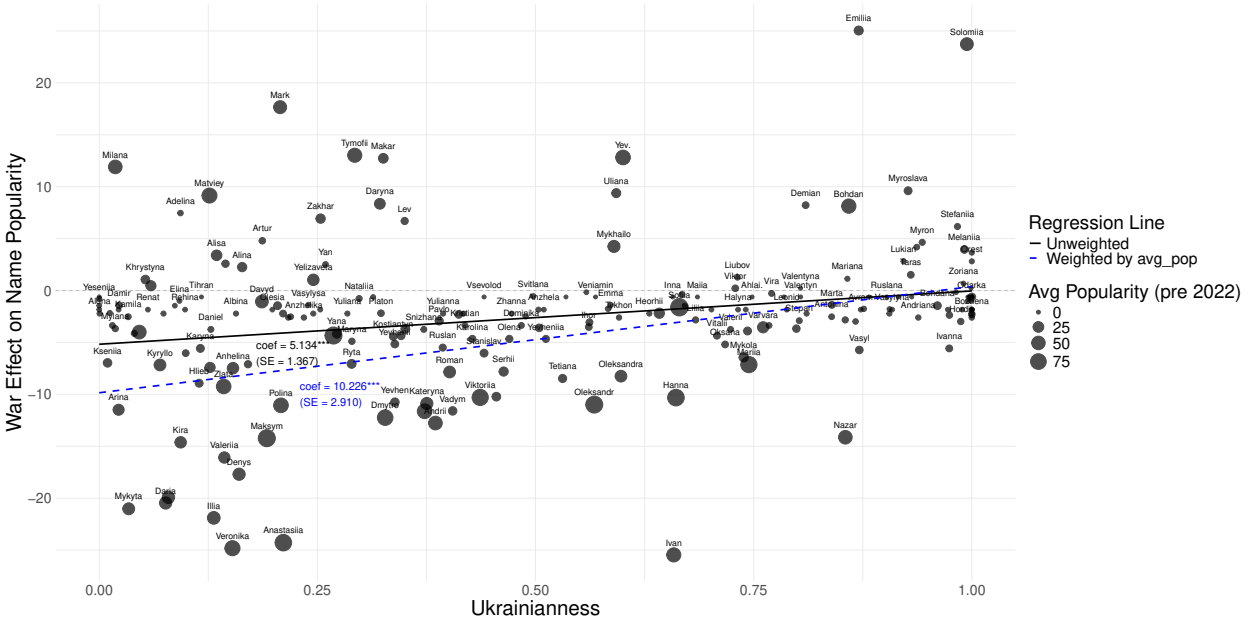


Figure C.4: Name-Level War Effects and the Ukrainianness Correlation

Notes: Each point represents a name. The vertical axis plots the estimated name-specific post–February 2022 effect $\hat{\tau}_n$ from equation (A19), which is obtained from a difference-in-differences specification with region–time and region–name fixed effects. The horizontal axis reports the name's prewar Ukrainianness score. Solid lines show the cross-sectional gradient from equation (A20), 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.

with these matched names, which account for approximately 73% of births to ethnically Ukrainian children over the sample period.

To test whether relatively more Ukrainian names became more common after the full-scale invasion, we estimate a specification analogous to Equation (6):

$$\text{Share}_{nto} = \sum_{s \neq 2021} \theta_s^{KZ} \left(\text{Ukrainianness}_n \times \mathbf{1}[t = s]\right) + \eta_{no} + \mu_{to} + \varepsilon_{nto} \tag{A21}$$

where Share_{nto} is the share of Ukrainian-nationality children born in region o and year t who receive name n, among all ethnically Ukrainian births. We include name–region fixed effects η_{no} and region–year fixed effects μ_{to}, so that θ_s^{KZ} reflects changes in the relative popularity of more- and less-Ukrainian names within the same region and year. Standard errors are two-way clustered by region and name.

Figure C.5 shows a clear post-2022 reallocation toward names with higher prewar Ukrainianness, paralleling the timing and direction of the pattern observed in Ukraine. By 2025, the relative gain in birth share for a fully Ukrainian vs. a fully Russian name was 0.20

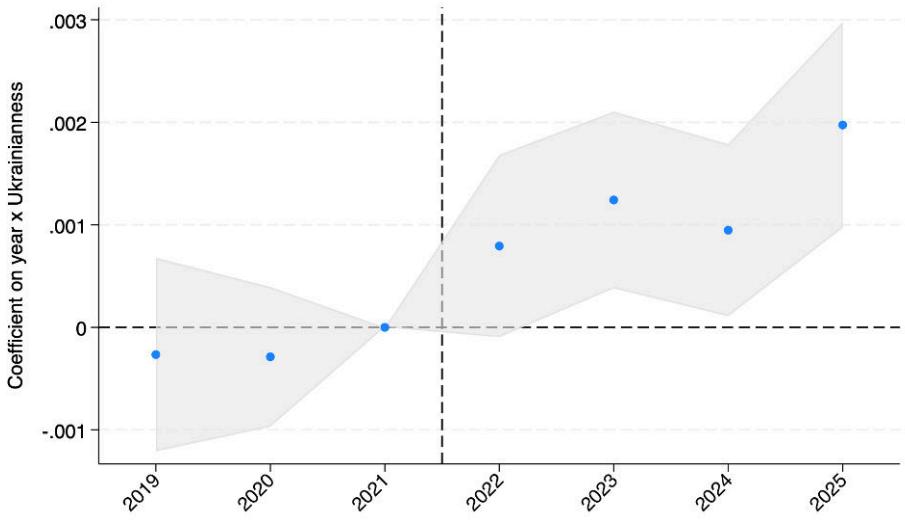


Figure C.5: Ukrainianness of Baby Names Among Ethnic Ukrainians in Kazakhstan

Notes: This figure depicts the evolution of baby-name popularity among ethnic Ukrainians in Kazakhstan as a function of names' prewar Ukrainianness. Specifically, it plots the coefficients θ_s^{KZ} from Equation (A21). The unit of observation is name $\times$ region $\times$ year. The outcome is the share of all ethnically Ukrainian births in region o and year t receiving name n. *Ukrainianness* is the same time-invariant score used in the main analysis, constructed from a name's prewar relative popularity in Ukraine vs. Russia. The specification includes name–region and region–year fixed effects; coefficients are normalized to 2021, the omitted year. Positive coefficients indicate that names with higher prewar Ukrainianness gained relative popularity among ethnic Ukrainians in Kazakhstan relative to 2021. Thus, an increase in the coefficients after 2022 indicates a shift toward more Ukrainian names following the Russian invasion of Ukraine. Standard errors are two-way clustered by region and name. Shaded areas denote 95% confidence intervals.

percentage points. This magnitude is equivalent to approximately 50% of the mean name share in 2021. While the coefficients are not directly comparable to those in our Ukraine analysis (because the outcomes differ), both exercises indicate sizable shifts toward more traditionally Ukrainian names following the invasion.

The Kazakhstan evidence is also informative for the interpretation of the main result. Because a similar shift is visible among ethnic Ukrainians living outside Ukraine, it is unlikely to be driven solely by Ukraine-specific institutional or economic incentives to adopt more Ukrainian-coded names. The result is also consistent with the status channel in our framework: changes in the relative appeal or status of Ukrainian and Russian identities can affect ethnic Ukrainians outside Ukraine as well. We do not interpret this exercise as isolating the status channel, however, because the war may also have increased the salience of Ukrainian–Russian cultural distinctions among Ukrainians abroad.

C.4 Music Consumption: Spotify Charts and Artist-Level Data

Scope and Main Sources. We measure music consumption using Spotify's weekly charts of the top-200 songs in Ukraine (Spotify, 2025a). Historical weekly chart data are scraped from the music-data tracking and aggregation website 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 also construct a list of Ukrainian artists using the Ukrainian-music platform NUAM Club (2025), Spotify (2025b), and Wikipedia (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. We then retrieve release dates and additional song metadata from Genius and the Spotify Web API. These release dates are used to define prewar and postwar 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 through the Genius API. After matching each charting track to a Genius entry, we classify the full lyric text with the `langdetect` Python library (version 1.0.9). Each song is assigned to its highest-probability language and grouped into four mutually exclusive categories: Ukrainian, Russian, Other (all remaining languages, mostly English), and Missing lyrics (no Genius entry, mostly tracks absent from the database and instrumentals). Missing lyrics account for 3.6% of unique charting tracks and 2.4% of streams; these songs are retained and tracked separately.

We validate the language classification in two ways. First, an independent classifier `langid` agrees with our labels for 98.7% of both unique tracks and chart entries. Second,

two native-speaking research assistants independently reviewed 100 randomly selected tracks, drawn in proportion to their chart presence, without seeing the automated labels. Manual review confirmed the algorithm's classification of the retrieved text in all 100 cases.

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).

Artist Classification by Attitude Toward the War. For the analysis by artists' political stance, we classify artists based on the Stars About War Project (2025). We match artists appearing in the Ukrainian Spotify charts to this source and assign available classifications (e.g., prowar, antiwar) 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 (2025), then we harmonize names, restricting the list to artists with at least one chart appearance in our sample. This procedure yields a list of 266 artists.

Artist Catalog Data and Release-Level Panel. For each artist in the Ukrainian-artist list, we retrieve the available catalog from Genius, 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 by Ukrainian artists.

Artist-Quarter Panel for Within-Artist Language Choice. To allow for within-artist language switching analysis, we aggregate the release-level data to an artist-by-quarter

panel. For each artist a, quarter t, and language l, we construct the release share

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

We then estimate specifications with artist and quarter fixed effects, identifying changes in language choice within the same artist over time.

C.5 Administrative Reform and Geographic Units

Ukraine's 2020 administrative reform consolidated the preexisting rayon system into a smaller number of larger districts, 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 postreform rayons. For simplicity, throughout the paper we refer to the pre-2020 rayons and cities of regional significance collectively as *rayons*.

In analyses of the 2014 war, we use fixed effects and clustering at the pre-2020 rayon level, corresponding to the administrative units governing respondents' local environment at the time. In analyses of the 2022 invasion, we instead use fixed effects and clustering at the post-2020 rayon level, consistent with the contemporaneous administrative structure.

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

C.6 Construction of Measures of Internally Displaced Persons

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

For 2014 through 2019, the ministry reported the stock of registered IDPs by oblast at irregular dates. We linearly interpolate these counts between reporting dates and set them to zero before September 2014. In April 2020, the ministry published the stock of IDPs

registered by the social-protection office, which corresponds to pre-2020 rayons and cities of significance. We allocate each oblast's stock at each date across its pre-2020 rayons in proportion to the rayons' shares of the oblast's registered IDPs in April 2020 and divide by rayon population. The resulting control measures registered IDPs per 100 residents of the pre-2020 rayon. Its within-oblast variation over time comes from changes in the oblast-level stock, while its cross-rayon distribution within each oblast is fixed at the April 2020 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 reduce the resulting overcount, we use administrative statistics from seven western oblasts—where incentives for cross-line registration were minimal—to benchmark the composition of genuine IDP registrations. Specifically, we compute in these oblasts the ratio of registered pensioners and persons with disabilities to all other registered IDPs. For each other oblast, we replace the observed number of pensioner and disabled registrants with the number implied by applying this ratio to its other registered IDPs. This normalization retains the number of such registrants expected under the western-oblast benchmark while removing registrations in excess of that benchmark.

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. The results are not sensitive to alternative constructions or to excluding IDP adjustments altogether.

D Additional Results

This appendix collects the additional tables and figures referenced in the main text, in order of first mention.

Table D.1: Ethnic Identity Before and After the 2014 War: Five-Category Gradient

	(1) Only Ukrainian	(2) More Ukrainian	(3) Mixed Ukr./Rus.	(4) More Russian	(5) Only Russian
Panel A: Postwar Shift					
Post–February 2014	7.127***	-0.447	-2.842***	-1.482***	-2.356***
	(1.051)	(0.326)	(0.768)	(0.312)	(0.319)
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: Difference-in-Differences					
Post–February 2014 × share of ethnic Russians	63.761***	-1.568	-27.952***	-12.431***	-21.810***
	(9.744)	(4.197)	(8.499)	(3.366)	(3.245)
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 examines changes in ethnic identity before and after the onset of the 2014 war. The dependent variables are indicators for five mutually exclusive ethnic-identity categories, ranging from Only Ukrainian to Only Russian, with three intermediate categories; all are multiplied by 100. 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 pre-2020 rayon level using the 2001 Census and varies from zero to one. All specifications include rayon fixed effects; Panel B specifications also include survey-wave fixed effects. Panel A omits wave fixed effects, because they would absorb the post–February 2014 indicator. All specifications control for sex, age, education, an urban/rural indicator, and rayon-level IDP inflows per population. Standard errors in parentheses are clustered at the rayon level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.2: Language Choice Before and After the 2022 Invasion: Three-Category Gradient

	(1) Replying in Ukrainian	(2) Replying in Mixed (Ukr.–Rus.)	(3) Replying in Russian
Panel A: Postwar 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: Difference-in-Differences			
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 examines changes in interview language before and after the 2022 invasion. The dependent variables are indicators for interviews conducted in Ukrainian, mixed Ukrainian–Russian, and Russian, respectively; all are multiplied by 100. 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 rayon level using the 2001 Census and varies from zero to one. All specifications include rayon fixed effects; Panel B specifications also include survey-wave fixed effects. Panel A specifications omit wave fixed effects, because they would absorb the post–February 2022 indicator. All specifications control for sex, age, education, an urban/rural indicator, the rayon-level share of ethnic Russians, and an indicator for being internally displaced after February 2022. Standard errors in parentheses are clustered at the rayon level. See Appendix C.5 for details on the geographic units. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: Nonlinearity in the Identity Changes Around the 2014 War and 2022 Invasion, by Quartile of the Prewar Share of Ethnic Russians

Panel A: 2014 War				
	Identified as Ukrainian Only (%)		Replying Not in Russian (%)	
	(1)	(2)	(3)	(4)
Post–February 2014 × 2nd quartile	4.261***	3.201**	-0.474	-1.833
	(1.364)	(1.415)	(1.474)	(1.539)
Post–February 2014 × 3rd quartile	8.044***	6.788***	0.351	-1.405
	(2.309)	(2.202)	(2.781)	(2.969)
Post–February 2014 × 4th quartile	21.718***	19.193***	3.817**	1.023
	(2.579)	(2.745)	(1.532)	(1.879)
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
Rayon FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓
Controls		✓		✓
Panel B: 2022 Invasion				
	Self-Identifying as Ukrainian (%)		Replying Not in Russian (%)	
	(1)	(2)	(3)	(4)
Post–February 2022 × 2nd quartile	1.836***	1.518***	8.895***	8.474***
	(0.491)	(0.493)	(1.285)	(1.271)
Post–February 2022 × 3rd quartile	3.862***	3.483***	22.223***	21.683***
	(0.953)	(0.906)	(1.545)	(1.507)
Post–February 2022 × 4th quartile	10.794***	10.074***	27.516***	25.661***
	(0.835)	(0.732)	(0.982)	(1.007)
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
Rayon FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓
Controls		✓		✓

Notes: This table examines nonlinearity in the effects of the 2014 war (Panel A) and the 2022 invasion (Panel B) by replacing the continuous share of ethnic Russians with indicators for its quartiles. The share is measured at the rayon level using the 2001 Census, and respondents are assigned to quartiles based on their rayon's share, from lowest (Q1) to highest (Q4). Each coefficient is the interaction of the relevant postconflict indicator with a quartile indicator. Q1 is omitted, so coefficients report differential changes relative to the lowest-share quartile. In Panel A, the dependent variable in columns (1) and (2) indicates exclusive ethnic Ukrainian identification; in Panel B, it indicates ethnic Ukrainian identification. In both panels, the dependent variable in columns (3) and (4) indicates that the interview was conducted in Ukrainian or a Ukrainian–Russian mix rather than in Russian. All outcomes are multiplied by 100. All specifications include survey-wave and rayon fixed effects (pre-2020 rayons in Panel A, post-2020 rayons in Panel B). Panel B specifications also include pre-2020 rayon quartile main effects, which are absorbed by the rayon fixed effects in Panel A. Even-numbered columns also include the controls used in the corresponding baseline specifications. Standard errors in parentheses are clustered at the rayon level. See Appendix C.5 for details on the geographic units. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4: Heterogeneity in the Effect of the 2014 War on Ethnic Identity

	Identified as Ukrainian Only (%)				
	Male	Age	Urban	Higher Education	South/ East
	(1)	(2)	(3)	(4)	(5)
Panel A: Postwar Shift and Individual/Regional Heterogeneity					
Post–February 2014	6.343***	8.833***	3.391***	6.042***	1.600***
	(1.088)	(1.579)	(1.075)	(1.005)	(0.545)
$\times X$	0.423	-0.044**	4.834***	1.855**	14.495***
	(0.450)	(0.019)	(1.843)	(0.863)	(1.846)
Dependent variable mean	79.757	79.757	79.757	79.782	79.757
Dependent variable st. dev.	40.181	40.181	40.181	40.163	40.181
Observations	87,751	87,751	87,751	87,272	87,751
IDP Control	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Panel B: Difference-in-Differences and Individual/Regional Heterogeneity					
Post–February 2014 $\times$ share of ethnic Russians	62.268***	88.653***	105.748***	61.808***	19.067*
	(10.858)	(13.948)	(24.083)	(10.203)	(11.002)
$\times X$	4.384	-0.482***	-45.094*	4.905	17.841
	(5.321)	(0.156)	(26.442)	(8.144)	(19.668)
Dependent variable mean	79.757	79.757	79.757	79.782	79.757
Dependent variable st. dev.	40.181	40.181	40.181	40.163	40.181
Observations	87,751	87,751	87,751	87,272	87,751
IDP Control	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓	✓
Post–February 2014 $\times X$	✓	✓	✓	✓	✓
Share of ethnic Russians $\times X$	✓	✓	✓	✓	

Notes: This table examines heterogeneity in the effect of the 2014 war on exclusive ethnic Ukrainian identification. The dependent variable equals one if the respondent identifies exclusively as Ukrainian and is multiplied by 100. Each column interacts the relevant postconflict variable with the individual or regional characteristic X listed in the column header. *Male* equals one for male respondents. *Age* is measured in years. *Urban* equals one for urban residents. *Higher Education* equals one for respondents with incomplete or complete higher education (levels 7–8) and zero for those with primary, secondary, or vocational education (levels 1–6). *South/East* equals one for respondents in the Southern or Eastern macroregions and zero for those in the Central or Western macroregions. Panel A reports coefficients on *Post–February 2014* and its interaction with X. Panel B reports coefficients on *Post–February 2014* $\times$ *share of ethnic Russians* and its interaction with X, where the share is measured at the rayon level using the 2001 Census and varies from zero to one. All specifications include rayon fixed effects and control for rayon-level IDP inflows per population; Panel B also includes survey-wave fixed effects. Each specification also includes the main effect of X (where not absorbed by the fixed effects); Panel B specifications also include the interactions of X with the post–February 2014 indicator and with the share of ethnic Russians (for *South/East*, the latter is absorbed by the rayon fixed effects), so that the coefficient on the triple interaction measures heterogeneity in the difference-in-differences gradient. Standard errors in parentheses are clustered at the rayon level. See Appendix C.5 for details on the geographic units. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5: Heterogeneity in the Effect of the 2014 War on Language Use

	Replying Not in Russian (%)				
	Male	Age	Urban	Higher Education	South/ East
	(1)	(2)	(3)	(4)	(5)
Panel A: Postwar Shift and Individual/Regional Heterogeneity					
Post–February 2014	0.642	2.086*	-0.052	0.602	0.164
	(1.339)	(1.217)	(1.100)	(1.201)	(1.382)
$\times X$	0.447	-0.028	1.541	1.526*	1.882
	(0.566)	(0.022)	(1.812)	(0.846)	(1.670)
Dependent variable mean	60.780	60.780	60.780	60.761	60.780
Dependent variable st. dev.	48.824	48.824	48.824	48.829	48.824
Observations	95,666	95,666	95,666	95,117	95,666
IDP Control	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Panel B: Difference-in-Differences and Individual/Regional Heterogeneity					
Post–February 2014 $\times$ share of ethnic Russians	9.290	18.518***	8.877	9.077	-20.544
	(5.935)	(6.673)	(19.138)	(5.808)	(26.222)
$\times X$	-0.585	-0.216**	-2.417	-2.758	35.044
	(2.868)	(0.108)	(19.923)	(5.837)	(27.309)
Dependent variable mean	60.780	60.780	60.780	60.761	60.780
Dependent variable st. dev.	48.824	48.824	48.824	48.829	48.824
Observations	95,666	95,666	95,666	95,117	95,666
IDP Control	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓	✓
Post–February 2014 $\times X$	✓	✓	✓	✓	✓
Share of ethnic Russians $\times X$	✓	✓	✓	✓	

Notes: This table explores heterogeneity in the baseline estimates for the effect of the 2014 war on language use in survey interviews. The dependent variable equals one if the interview was conducted in a language other than Russian (Ukrainian or mixed Ukrainian–Russian), multiplied by 100 (percentage points). Each column presents a specification that interacts the main variable of interest with the individual or regional characteristic X listed in the column header. *Male* equals one for male respondents. *Urban* equals one for urban residents. *Higher Education* equals one for respondents with incomplete or complete higher education (levels 7–8) and zero for those with primary, secondary, or vocational education (levels 1–6). *South/East* equals one for respondents in Eastern or Southern macroregions, zero for Central or Western. Panel A reports coefficients on *Post–February 2014* and its interaction $\times X$ with the column characteristic. Panel B reports coefficients on *Post–February 2014* $\times$ *Share of ethnic Russians* and its interaction $\times X$, where the share is measured at the rayon level using the 2001 Census. All specifications control for rayon-level IDP inflows per population and include rayon fixed effects. Panel B also includes survey-wave fixed effects and the interactions of X with the post–February 2014 indicator and with the share of ethnic Russians (for *South/East*, the latter is absorbed by the rayon fixed effects), so that the coefficient on the triple interaction measures heterogeneity in the difference-in-differences gradient. Standard errors in parentheses are clustered at the rayon level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.6: Heterogeneity in the Effect of the 2022 Invasion on Ethnic Identity

	Identified as Ukrainian (%)				
	Male	Age	Urban	Higher Education	South/ East
	(1)	(2)	(3)	(4)	(5)
Panel A: Postwar Shift and Individual/Regional Heterogeneity					
Post–February 2022	5.464***	-0.232	3.187***	5.185***	2.800***
	(0.790)	(0.672)	(0.511)	(0.613)	(0.290)
$\times X$	-0.559	0.118***	2.533***	-0.020	6.737***
	(0.495)	(0.013)	(0.710)	(0.440)	(0.752)
Dependent variable mean	92.254	92.254	92.254	92.251	92.254
Dependent variable st. dev.	26.732	26.732	26.732	26.736	26.732
Observations	50,673	50,673	50,673	50,512	50,673
Controls	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Panel B: Difference-in-Differences and Individual/Regional Heterogeneity					
Post–February 2022 $\times$ share of ethnic Russians	42.412***	15.027**	55.718***	36.910***	28.227***
	(3.479)	(6.349)	(8.266)	(3.343)	(7.257)
$\times X$	-7.073	0.502***	-15.903*	5.285	5.922
	(5.291)	(0.147)	(8.895)	(5.642)	(8.817)
Dependent variable mean	92.254	92.254	92.254	92.251	92.254
Dependent variable st. dev.	26.732	26.732	26.732	26.736	26.732
Observations	50,673	50,673	50,673	50,512	50,673
Controls	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓	✓
Post–February 2022 $\times X$	✓	✓	✓	✓	✓
Share of ethnic Russians $\times X$	✓	✓	✓	✓	✓

Notes: This table explores heterogeneity in the baseline estimates for the effect of the 2022 invasion on the probability of self-identifying as Ukrainian. The dependent variable equals one if the respondent identifies as Ukrainian, multiplied by 100 (percentage points). Each column presents a specification that interacts the main variable of interest with the individual or regional characteristic X listed in the column header. *Male* equals one for male respondents. *Urban* equals one for urban residents. *Higher Education* equals one for respondents with incomplete or complete higher education (levels 7–8) and zero for those with primary, secondary, or vocational education (levels 1–6). *South/East* equals one for respondents in Eastern or Southern macroregions, zero for Central or Western. Panel A reports coefficients on *Post–February 2022* and its interaction $\times X$ with the column characteristic. Panel B reports coefficients on *Post–February 2022* $\times$ *Share of Ethnic Russians* and its interaction $\times X$, where the share is measured at the rayon level using the 2001 Census. All specifications control for post-2022 displacement status and include post-2020 rayon fixed effects and the pre-2020 rayon share of ethnic Russians. Panel B also includes survey-wave fixed effects and the interactions of X with the post–February 2022 indicator and with the share of ethnic Russians, so that the coefficient on the triple interaction measures heterogeneity in the difference-in-differences gradient. Standard errors in parentheses are clustered at the rayon level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.7: Heterogeneity in the Effect of the 2022 Invasion on Language Use

	Replying Not in Russian (%)				
	Male	Age	Urban	Higher Education	South/ East
	(1)	(2)	(3)	(4)	(5)
Panel A: Postwar Shift and Individual/Regional Heterogeneity					
Post–February 2022	19.780***	28.735***	9.153***	16.467***	14.192***
	(1.857)	(3.369)	(1.137)	(1.513)	(2.715)
$\times X$	-0.471	-0.183***	13.106***	6.451***	15.019***
	(0.719)	(0.035)	(1.487)	(1.005)	(3.000)
Dependent variable mean	69.675	69.675	69.675	69.652	69.675
Dependent variable st. dev.	45.967	45.967	45.967	45.977	45.967
Observations	48,976	48,976	48,976	48,744	48,976
Controls	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Panel B: Difference-in-Differences and Individual/Regional Heterogeneity					
Post–February 2022 $\times$ share of ethnic Russians	91.349***	153.000***	119.397***	78.777***	157.040***
	(10.371)	(18.436)	(18.582)	(7.646)	(24.954)
$\times X$	-6.220	-1.293***	-41.860**	15.256*	-113.145***
	(7.498)	(0.221)	(19.670)	(8.472)	(26.950)
Dependent variable mean	69.675	69.675	69.675	69.652	69.675
Dependent variable st. dev.	45.967	45.967	45.967	45.977	45.967
Observations	48,976	48,976	48,976	48,744	48,976
Controls	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓	✓
Post–February 2022 $\times X$	✓	✓	✓	✓	✓
Share of ethnic Russians $\times X$	✓	✓	✓	✓	✓

Notes: This table explores heterogeneity in the baseline estimates for the effect of the 2022 invasion on language use in survey interviews. The dependent variable equals one if the interview was conducted in a language other than Russian (Ukrainian or mixed Ukrainian–Russian), multiplied by 100 (percentage points). Each column presents a specification that interacts the main variable of interest with the individual or regional characteristic X listed in the column header. *Male* equals one for male respondents. *Urban* equals one for urban residents. *Higher Education* equals one for respondents with incomplete or complete higher education (levels 7–8) and zero for those with primary, secondary, or vocational education (levels 1–6). *South/East* equals one for respondents in Eastern or Southern macroregions, zero for Central or Western. Panel A reports coefficients on *Post–February 2022* and its interaction $\times X$ with the column characteristic. Panel B reports coefficients on *Post–February 2022* $\times$ *Share of ethnic Russians* and its interaction $\times X$, where the share is measured at the rayon level using the 2001 Census. All specifications control for post-2022 displacement status and include post-2020 rayon fixed effects and the pre-2020 rayon share of ethnic Russians. Panel B also includes survey-wave fixed effects and the interactions of X with the post–February 2022 indicator and with the share of ethnic Russians, so that the coefficient on the triple interaction measures heterogeneity in the difference-in-differences gradient. Standard errors in parentheses are clustered at the rayon level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.8: Interviewer-by-Wave Fixed Effects: The 2014 War

	Identified as Ukrainian Only (%)		Replying Not in Russian (%)	
	(1)	(2)	(3)	(4)
Post–February 2014 × share of ethnic Russians	57.756***	49.931***	4.378	-6.577
	(8.923)	(8.947)	(5.514)	(6.106)
Dependent variable mean	79.489	79.514	60.664	60.649
Dependent variable st. dev.	40.379	40.361	48.850	48.853
Observations	83,496	83,045	91,264	90,747
Controls		✓		✓
Rayon FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓
Interviewer × wave FE	✓	✓	✓	✓

Notes: This table re-estimates Panel B of Table 1, adding interviewer-by-wave fixed effects. The sample includes the 51 survey waves between 2008 and 2019 with interviewer identifiers; November 2013 and September 2018 waves carry no interviewer identifiers and are excluded. Interviewer identifiers cannot be linked across waves. Dependent variables are indicators multiplied by 100: in columns (1) and (2), the respondent identifies exclusively as Ukrainian rather than as Russian or mixed; in columns (3) and (4), the respondent replies in Ukrainian or mixed Ukrainian–Russian rather than exclusively in Russian. Each column reports the coefficient on *Post–February 2014 × share of ethnic Russians*, with the share measured at the rayon level in the 2001 Census on a zero-to-one scale. All specifications include rayon, survey-wave, and interviewer-by-wave fixed effects. Columns (2) and (4) also control for sex, age, education, urban/rural residence, and rayon-level IDP inflows per population. Interviewer identifiers are unavailable for the 2020–2024 period, preventing an analogous check for the 2022 invasion. Interviewer characteristics are also unavailable, precluding heterogeneity analysis along this dimension. Standard errors in parentheses are clustered at the rayon level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.9: Identity Shifts in Areas Far From the Ceiling

Panel A: 2014 War

	Identified as Ukrainian Only (%)		Replying Not in Russian (%)	
	South & East (1)	Above-Median Russian Share (2)	South & East (3)	Above-Median Russian Share (4)
Post–February 2014 × share of ethnic Russians	33.639**	66.432***	15.170	19.011
	(15.886)	(19.422)	(12.021)	(16.694)
Dependent variable mean	61.151	65.748	24.043	28.164
Dependent variable st. dev.	48.741	47.456	42.735	44.981
Observations	34,171	39,732	37,770	43,665
Controls	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓

Panel B: 2022 Invasion

	Identified as Ukrainian (%)		Replying Not in Russian (%)	
	South & East (1)	Above-Median Russian Share (2)	South & East (3)	Above-Median Russian Share (4)
Post–February 2022 × share of ethnic Russians	35.319***	42.384***	43.728***	23.048***
	(4.738)	(6.487)	(10.373)	(8.277)
Dependent variable mean	86.929	88.292	47.519	51.090
Dependent variable st. dev.	33.709	32.153	49.940	49.989
Observations	20,198	24,777	19,651	24,046
Controls	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓

Notes: This table examines whether the baseline difference-in-differences estimates reflect ceiling effects: Ukrainian identification was already nearly universal in rayons with few ethnic Russians, leaving them little scope to shift further and potentially exaggerating the estimated gradient. Panel A re-estimates the baseline specification for the 2014 war; Panel B does the same for the 2022 invasion. Columns (1) and (3) restrict the sample to the Southern and Eastern macroregions; columns (2) and (4) restrict it to rayons with above-median prewar shares of ethnic Russians. In Panel A, the dependent variable in columns (1) and (2) indicates exclusive ethnic Ukrainian identification; in Panel B, it indicates ethnic Ukrainian identification. In both panels, the dependent variable in columns (3) and (4) indicates that the interview was conducted in Ukrainian or a Ukrainian–Russian mix rather than in Russian. All outcomes are multiplied by 100. Panel A reports coefficients on *Post–February 2014 × share of ethnic Russians* and Panel B on *Post–February 2022 × share of ethnic Russians*, where the share is measured at the pre-2020 rayon level using the 2001 Census and varies from zero to one. All specifications include survey-wave and rayon fixed effects (pre-2020 rayons in Panel A, post-2020 rayons in Panel B). Panel A specifications control for sex, age, education, an urban/rural indicator, and rayon-level IDP inflows per population. Panel B specifications control for sex, age, education, an urban/rural indicator, the pre-2020 rayon-level share of ethnic Russians, and post-2022 displacement status. Standard errors in parentheses are clustered at the rayon level. See Appendix C.5 for details on the geographic units. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.10: Google Search Interest in Spotify by Oblast Prior to the 2022 Invasion

Region	Google Search Interest in Spotify
Kyiv city	100
Kharkiv oblast	97
Lviv oblast	90
Odesa oblast	73
Dnipropetrovsk oblast	66
Cherkasy oblast	66
Khmelnytskyi oblast	65
Zaporizhzhia oblast	65
Chernivtsi oblast	64
Ivano-Frankivsk oblast	64
Rivne oblast	63
Kherson oblast	62
Zakarpattia oblast	62
Mykolaiv oblast	62
Vinnytsia oblast	60
Poltava oblast	60
Kirovohrad oblast	59
Chernihiv oblast	58
Ternopil oblast	58
Kyiv oblast	58
Volyn oblast	58
Zhytomyr oblast	55
Sumy oblast	54
Luhansk oblast	52
Donetsk oblast	52

Notes: This table reports Google Trends search interest in Spotify by oblast between July 19, 2020, when Spotify entered the Ukrainian market, and February 19, 2022. Values are normalized on a 0–100 scale within Ukraine, where 100 corresponds to the oblast with the highest relative search intensity over the period. Crimea and Sevastopol, occupied since 2014, are excluded.

Table D.11: Baby-Name Popularity and Ukrainianness

	Popular Name Indicator					
	(1)	(2)	(3)	(4)	(5)	(6)
Sample	All	All	Boys	Girls	All	All
Heterogeneity					List Length	Russian Share
Post–February 2022 × Ukrainianness	5.016**	5.148**	4.050	6.130***	1.774	3.335
	(1.849)	(1.874)	(2.654)	(2.124)	(1.431)	(2.313)
Post–February 2022 × Ukrainianness × long name list					6.393**	
					(2.410)	
Post–February 2022 × Ukrainianness × share of ethnic Russians						18.233
						(12.445)
Dependent variable mean	13.593	13.593	14.774	12.528	13.593	13.593
Dependent variable st. dev.	34.271	34.271	35.485	33.104	34.271	34.271
Observations	83,134	83,134	39,400	43,734	83,134	83,134
Oblast × Period FE		✓	✓	✓	✓	✓
Oblast × Name FE		✓	✓	✓	✓	✓

Notes: This table examines the post–February 2022 change in the association between a baby name's popularity and its prewar Ukrainianness. The unit of observation is name × oblast × period, where periods are years before 2018 and half-years thereafter. The dependent variable equals one if a name appears in the Ministry of Justice list of most popular baby names in an oblast-period, multiplied by 100. *Ukrainianness* is the time-invariant name-level score described in Appendix C.3, constructed from the name's prewar relative popularity in Ukraine vs. Russia, and varies from zero to one. All columns report the coefficient on *Post–February 2022 × Ukrainianness*: columns (1) and (2) use the full sample, column (3) boys' names, and column (4) girls' names. Column (1) includes no fixed effects; columns (2)–(6) include oblast–period and oblast–name fixed effects. Column (5) adds the interaction with an indicator for above-median popularity-list length; column (6) adds the interaction with the oblast-level share of ethnic Russians measured using the 2001 Census. Standard errors in parentheses are two-way clustered by oblast and name. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.12: Local Exposure to Violent War Events and Russian-Language Replies

	Nonoccupied Rayons				*Never-Occupied Rayons*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Hospital Events	Civilian Casualty	Military Casualty	Air/ Artillery/ Drone	Hospital Events	Civilian Casualty	Military Casualty	Air/ Artillery/ Drone
Hospital events	-1853.278**				-3442.988			
	(720.354)				(3379.211)			
Civilian casualty events		-136.899**				-464.845*		
		(62.493)				(239.523)		
Military casualty events			-548.897**				-1909.367***	
			(247.791)				(651.080)	
Air/artillery/drone events				-63.181***				-144.599**
				(23.757)				(61.319)
Dependent variable mean	0.265	0.265	0.265	0.265	0.232	0.232	0.232	0.232
Dependent variable st. dev.	0.442	0.442	0.442	0.442	0.422	0.422	0.422	0.422
Observations	40,781	40,781	40,781	40,781	30,054	30,054	30,054	30,054
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Rayon FE	✓	✓	✓	✓	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table examines the association between local exposure to violent war events and interview language, following Equation (8), in the survey waves used in our 2020–2024 analysis. The dependent variable equals one if the interview was conducted in Russian and is not multiplied by 100. Each column reports the coefficient on the indicated type of VIINA event (Zhukov, 2022): hospital events, civilian-casualty events, military-casualty events, or air, artillery, and drone attacks. Exposure is measured as the number of events occurring since the preceding survey wave in the respondent's preinvasion rayon of residence, divided by its 2021 population; it is set to zero before February 24, 2022. The sample excludes respondents displaced after February 24, 2022 (3.4% of the nonoccupied sample). Columns (1)–(4) use nonoccupied rayons, excluding Donetsk, Luhansk, Zaporizhzhia, and Kherson oblasts, Crimea, and Sevastopol; columns (5)–(8) use never-occupied rayons, also excluding Kyiv, Mykolaiv, Sumy, Kharkiv, and Chernihiv oblasts (the city of Kyiv remains). All specifications include post-2020-rayon and survey-wave fixed effects, the rayon-level 2001 Census share of ethnic Russians interacted with survey-wave indicators, and controls for age, sex, education, urban/rural residence, and self-assessed family finances. Standard errors in parentheses are clustered at the rayon level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.13: Ethnic and Linguistic Identity Before and After the 2014 War: Occupied Donbas

	Identified as Ukrainian Only (%)		Replying Not in Russian (%)	
	(1)	(2)	(3)	(4)
Panel A: Postwar Shift and Occupied Donbas Heterogeneity				
Post–February 2014	8.439***	7.199***	1.655	0.637
	(1.125)	(1.016)	(1.663)	(1.659)
Post–February 2014 × occupied Donbas	-24.233***	-23.565***	-0.312	1.045
	(4.074)	(3.873)	(2.859)	(2.534)
Dependent variable mean	72.357	72.374	54.931	54.906
Dependent variable st. dev.	44.723	44.715	49.757	49.759
Observations	57,006	56,613	64,235	63,796
Rayon FE	✓	✓	✓	✓
Controls		✓		✓
Panel B: Difference-in-Differences and Occupied Donbas Heterogeneity				
Post–February 2014 × share of ethnic Russians	37.316***	31.860***	11.080*	3.081
	(9.747)	(9.603)	(5.682)	(5.445)
Post–February 2014 × occupied Donbas	-34.732***	-32.852***	-3.665	0.176
	(5.181)	(5.003)	(3.791)	(3.542)
Dependent variable mean	72.357	72.374	54.931	54.906
Dependent variable st. dev.	44.723	44.715	49.757	49.759
Observations	57,006	56,613	64,235	63,796
Rayon FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓
Controls		✓		✓

Notes: This table examines whether identity consolidation after the 2014 war differs between rayons that came under Russian control and the rest of Ukraine. The sample is restricted to survey waves through 2015, the last year in which KIIS surveyed the occupied areas of Donbas; Crimea is excluded throughout. In columns (1) and (2), the dependent variable equals one if the respondent identifies exclusively as Ukrainian; in columns (3) and (4), it equals one if the interview was conducted in a language other than Russian (Ukrainian or mixed Ukrainian–Russian). Both are multiplied by 100. *Occupied Donbas* is an indicator equal to one for rayons that fell under Russian control after February 2014. Panel A reports coefficients on *Post–February 2014* and *Post–February 2014 × occupied Donbas*. Panel B adds *Post–February 2014 × share of ethnic Russians*, retaining the occupied-Donbas interaction; the share is measured at the rayon level using the 2001 Census and varies from zero to one. All specifications include rayon fixed effects; Panel B specifications also include survey-wave fixed effects. Panel A omits wave fixed effects because they would absorb the post–February 2014 indicator. Columns (2) and (4) also control for sex, age, education, an urban/rural indicator, and rayon-level IDP inflows per population. Standard errors in parentheses are clustered at the rayon level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

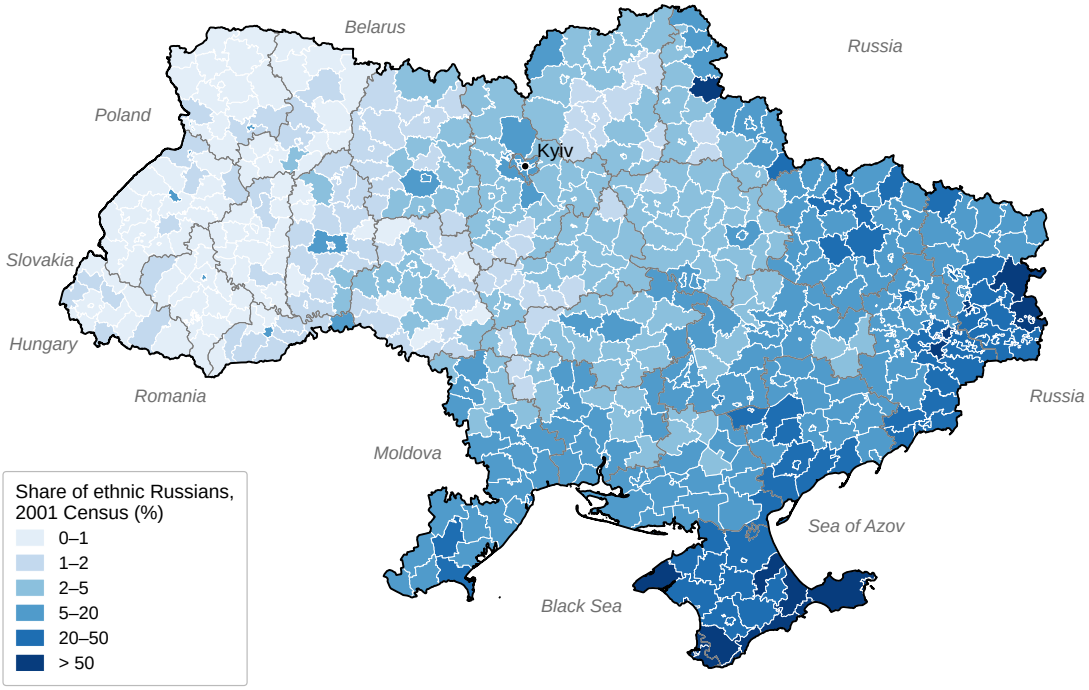


Figure D.1: Share of Ethnic Russians by Rayon, 2001 Census

Notes: This map shows the rayon-level share of residents reporting Russian ethnicity in the 2001 Census, our main measure of prewar ethnolinguistic proximity to Russia. Units are pre-2020 rayons and cities of regional significance, drawn from GADM (version 2.5) boundaries; thin gray lines mark oblast borders.

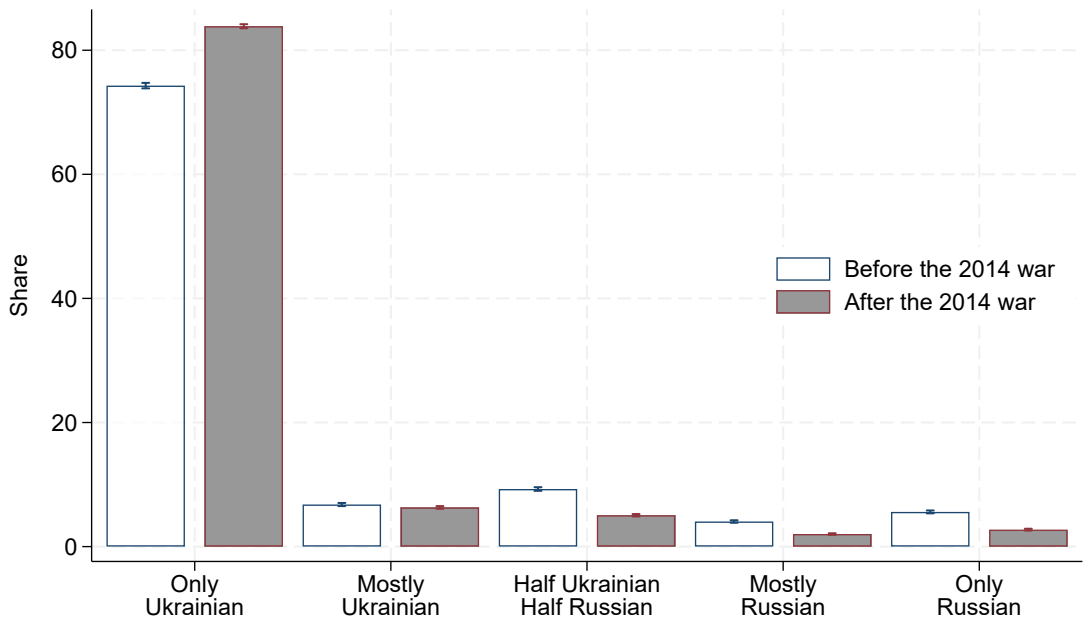


Figure D.2: Reported Ethnic Identity Before and After the 2014 War

Notes: This figure plots the mean percentage of KIIS respondents in each of the five categories of the ethnic-identity scale before the 2014 war (June 2008–February 2014) and after it (March 2014–February 2020). The sample excludes Crimea and rayons fully or partially occupied after 2014. Error bars denote 95% confidence intervals.

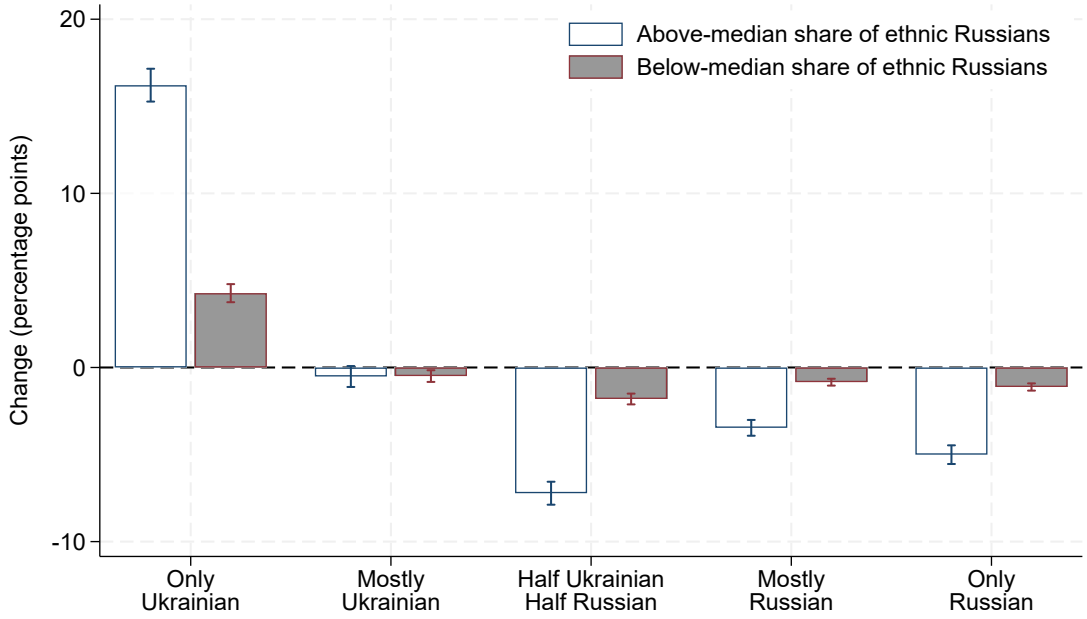


Figure D.3: Change in Ethnic Identity After the 2014 War, by Prewar Ethnic Composition

Notes: This figure plots the change in the mean percentage of KIIS respondents in each category of the ethnic-identity scale between the period after the 2014 war (March 2014–February 2020) and the period before it (June 2008–February 2014): $\Delta\bar{x} = \bar{x}_{\text{Mar 2014–Feb 2020}} - \bar{x}_{\text{Jun 2008–Feb 2014}}$. White bars are respondents in rayons with an above-median prewar share of ethnic Russians, measured using the 2001 Census; gray bars are respondents in rayons with a below-median share. The sample excludes Crimea and rayons fully or partially occupied after 2014. Error bars denote 95% confidence intervals.

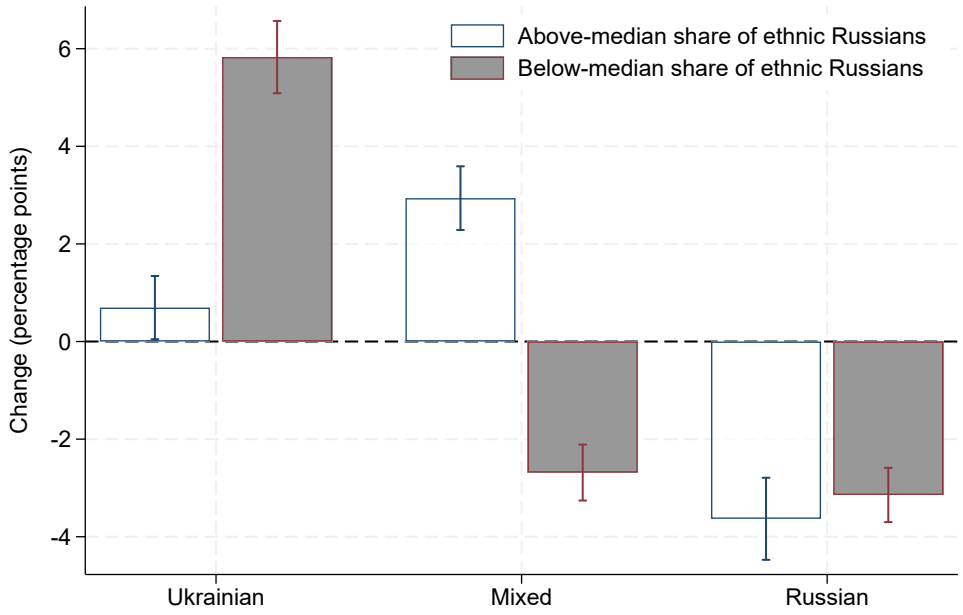


Figure D.4: Change in Reply Language After the 2014 War, by Prewar Ethnic Composition

Notes: This figure plots the change in the mean percentage of KIIS respondents replying in Ukrainian, in mixed Ukrainian–Russian, and in Russian between the period after the 2014 war (March 2014–February 2020) and the period before it (June 2008–February 2014): $\Delta\bar{x} = \bar{x}_{\text{Mar 2014–Feb 2020}} - \bar{x}_{\text{Jun 2008–Feb 2014}}$. White bars are respondents in rayons with an above-median prewar share of ethnic Russians, measured using the 2001 Census; gray bars are respondents in rayons with a below-median share. The sample excludes Crimea and rayons fully or partially occupied after 2014. Error bars denote 95% confidence intervals.

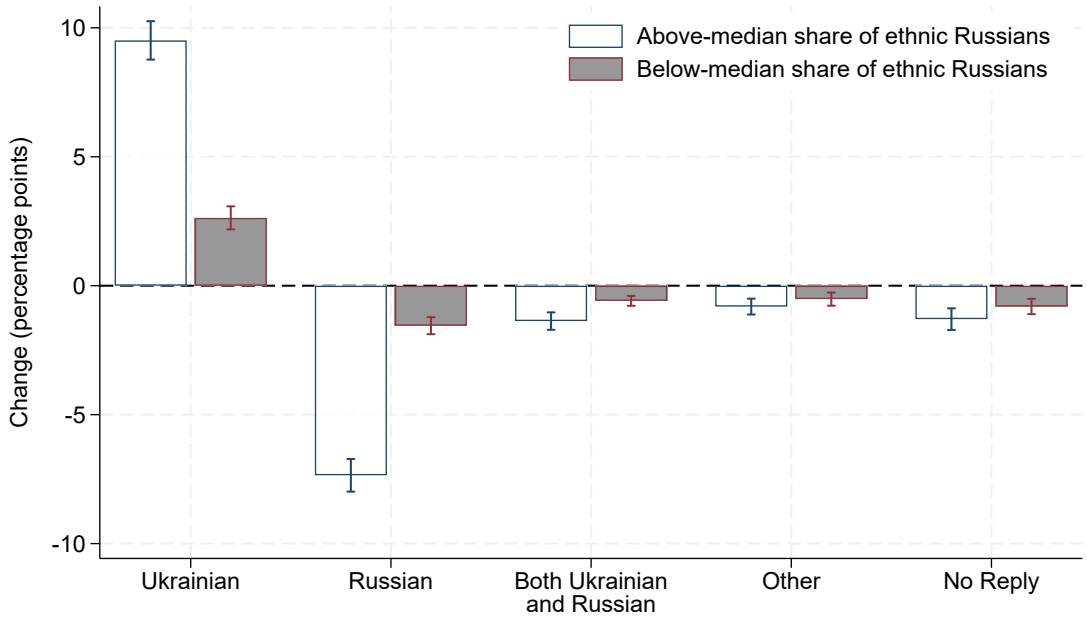


Figure D.5: Change in Ethnic Identity Post 2022 Invasion, by Prewar Ethnic Composition

Notes: This figure plots the change in the mean percentage of KIIS respondents naming Ukrainian, Russian, both Ukrainian and Russian, or another ethnicity in the open-ended nationality question, or not replying, between the period after Russia's full-scale invasion (March 2022–February 2024) and the period before it (April 2020–February 2022): $\Delta\bar{x} = \bar{x}_{\text{Mar 2022–Feb 2024}} - \bar{x}_{\text{Apr 2020–Feb 2022}}$. White bars are respondents in pre-2020 rayons with an above-median prewar share of ethnic Russians, measured using the 2001 Census; gray bars are respondents in rayons with a below-median share. Error bars denote 95% confidence intervals.

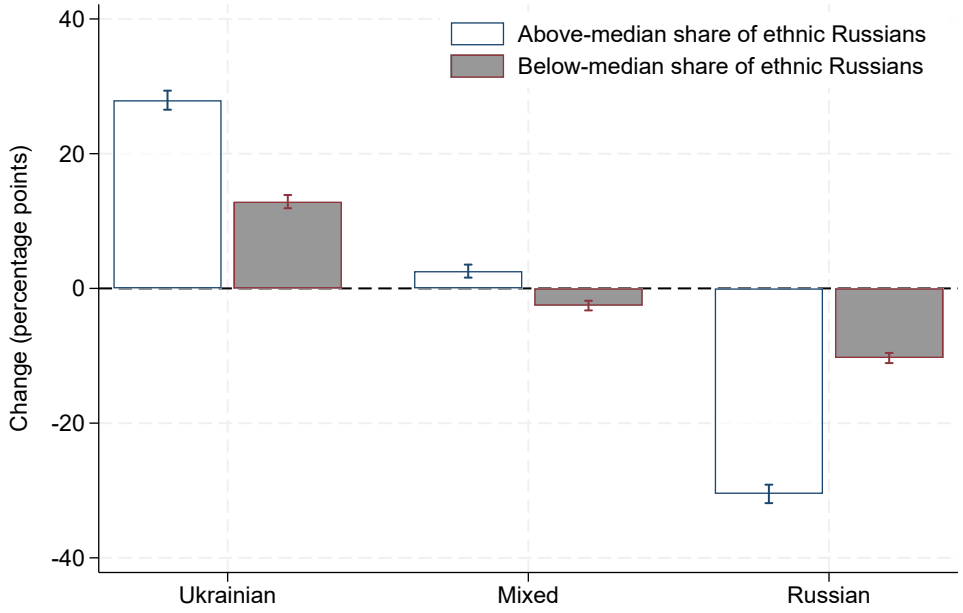
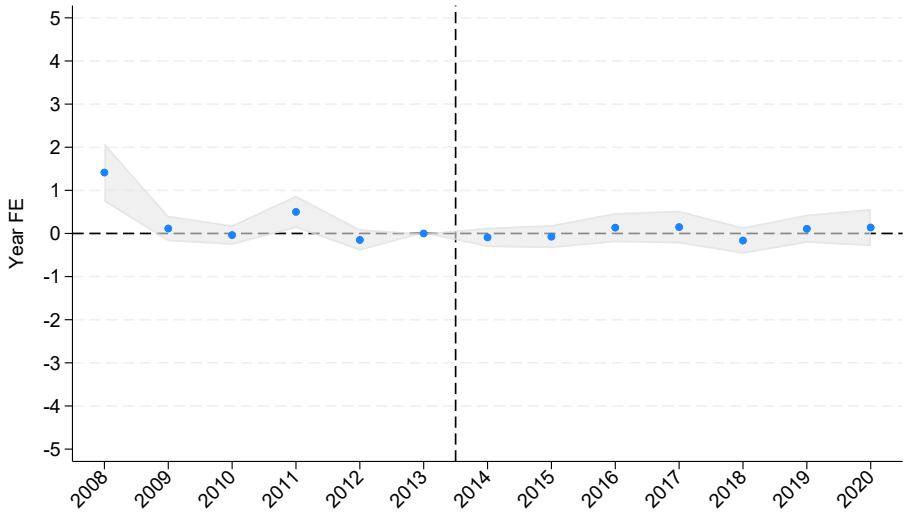
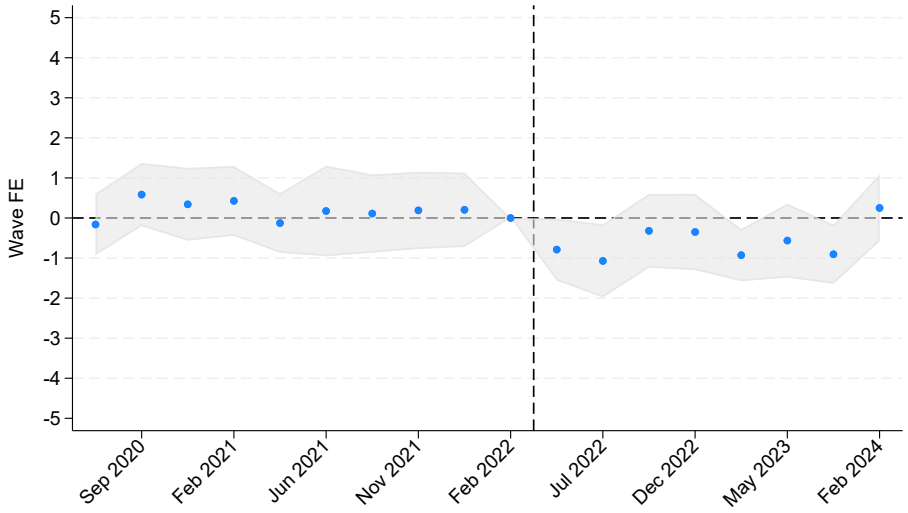


Figure D.6: Change in Reply Language Post 2022 Invasion, by Prewar Ethnic Composition

Notes: The figure plots the change in the mean percentage of KIIS respondents replying in Ukrainian, in mixed Ukrainian–Russian, and in Russian between the period after Russia's full-scale invasion (March 2022–February 2024) and the period before it (April 2020–February 2022): $\Delta\bar{x} = \bar{x}_{\text{Mar 2022–Feb 2024}} - \bar{x}_{\text{Apr 2020–Feb 2022}}$. White bars are respondents in pre-2020 rayons with an above-median prewar share of ethnic Russians, measured using the 2001 Census; gray bars are respondents in rayons with a below-median share. The sample covers the 2020–2024 survey waves and excludes Donetsk, Luhansk, Kherson, and Zaporizhzhia oblasts. Error bars denote 95% confidence intervals.



(a) Year Fixed Effects Around the 2014 War



(b) Survey-Wave Fixed Effects Around the 2022 Invasion

Figure D.7: No-Reply Rate on the Ethnicity Question: Time Fixed Effects

Notes: This figure plots time fixed effects from regressions of an indicator for not answering the ethnicity question, multiplied by 100, on period indicators, rayon fixed effects, and controls. Panel A uses the 2008–2020 waves, with year indicators normalized to 2013, pre-2020 rayon fixed effects, and controls for sex, age, education, and an urban/rural indicator; the sample excludes Crimea and rayons fully or partially occupied after 2014. Panel B uses the 2020–2024 waves, with survey-wave indicators normalized to the February 2022 wave, the last preinvasion wave, post-2020 rayon fixed effects, and controls for sex, age, education, an urban/rural indicator, the share of ethnic Russians in the respondent's pre-2020 rayon, and an indicator for being internally displaced. Vertical dashed lines mark the onset of hostilities in 2014 and Russia's full-scale invasion in 2022. Standard errors are clustered at the rayon level. Shaded areas denote 95% confidence intervals.

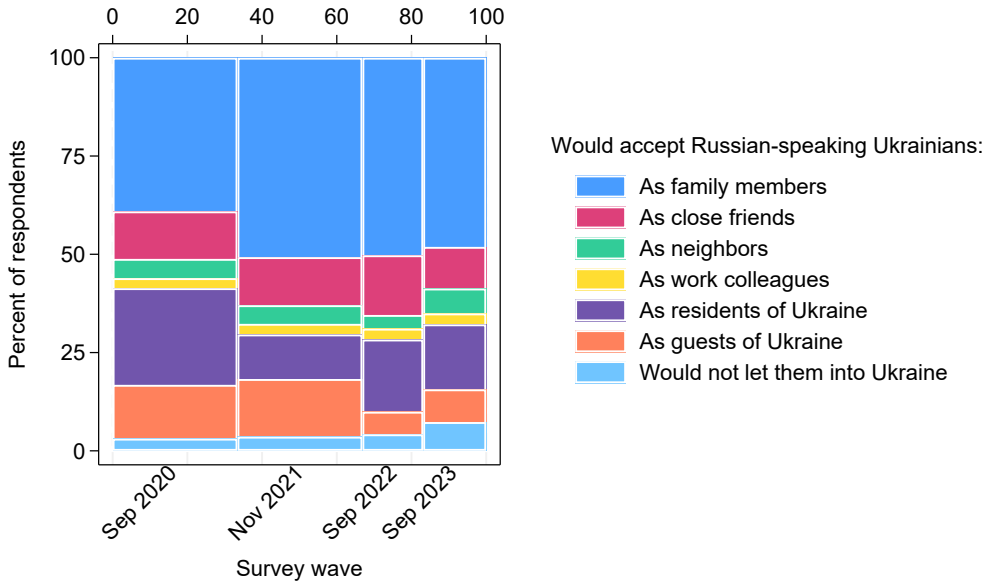


Figure D.8: Attitudes Toward Russian-Speaking Ukrainians

Notes: This figure plots the distribution of answers to the KIIS question asking about the closest relationship with Russian-speaking Ukrainians that the respondent would accept, from accepting them as family members to not letting them into Ukraine, in the four waves that asked the question (September 2020, November 2021, September 2022, and September 2023). Each bar sums to 100% within a wave; bar widths are proportional to the number of respondents in the wave.

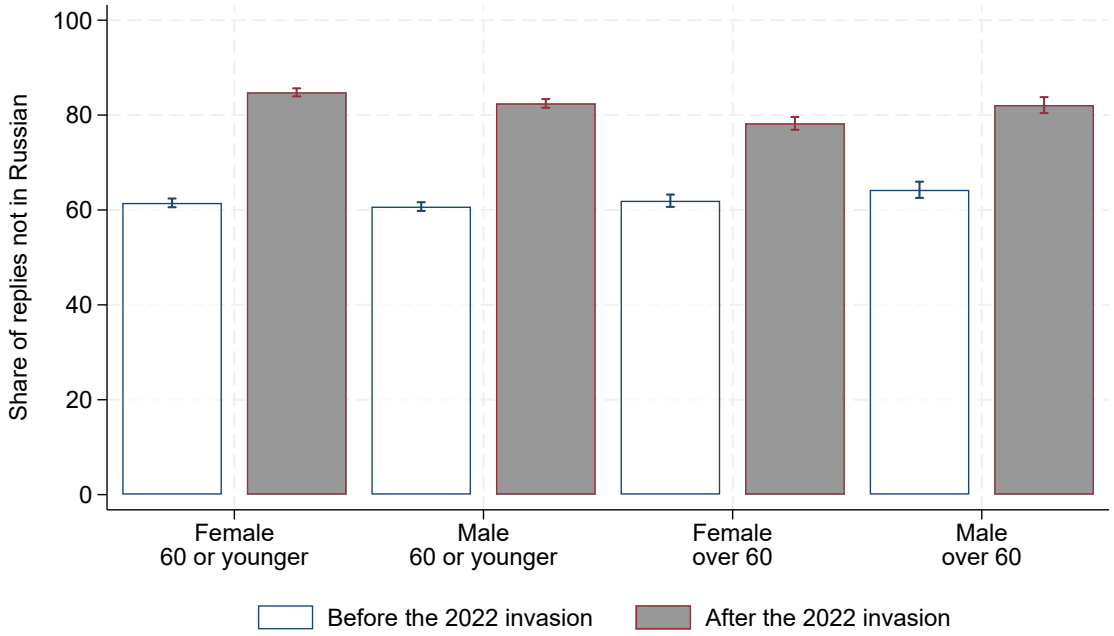


Figure D.9: Change in Language Use in Survey Interviews by Sex and Age

Notes: This figure examines whether the postinvasion shift away from Russian in KIIS survey interviews differs by sex and by whether respondents are above age 60. It reports the share of respondents replying in a language other than Russian (Ukrainian or a Ukrainian–Russian mix) before Russia's full-scale invasion (April 2020–February 2022) and after it (March 2022–February 2024), separately by sex and age group (60 or younger vs. over 60). Error bars denote 95% confidence intervals.

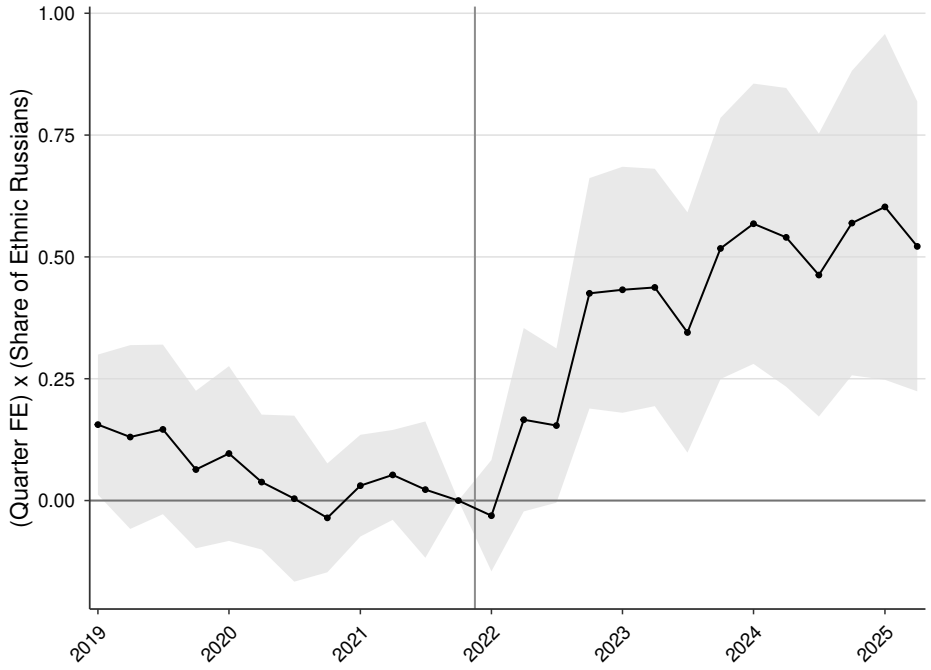


Figure D.10: Ukrainian Search Spelling by Prewar Ethnic Russian Share

Notes: This figure examines whether the postinvasion shift toward Ukrainian-spelling Google searches was larger in oblasts with higher prewar shares of ethnic Russians. The plotted coefficients are interactions between quarter indicators and the oblast's ethnic Russian share in the 2001 Census. The outcome is the Ukrainian-spelling share of searches within each Ukrainian–Russian word pair. The specification includes oblast–word-pair and quarter fixed effects, and coefficients are normalized to 2021Q4. Positive coefficients indicate a relatively larger shift toward Ukrainian spelling in oblasts with higher prewar ethnic Russian shares. The sample covers 21 oblasts from 2019Q1 onward, excluding Crimea, Sevastopol, and the four oblasts Russia claimed to annex in 2022. Standard errors are two-way clustered by oblast and word pair; shaded areas denote 95% confidence intervals. The vertical line marks Russia's full-scale invasion.

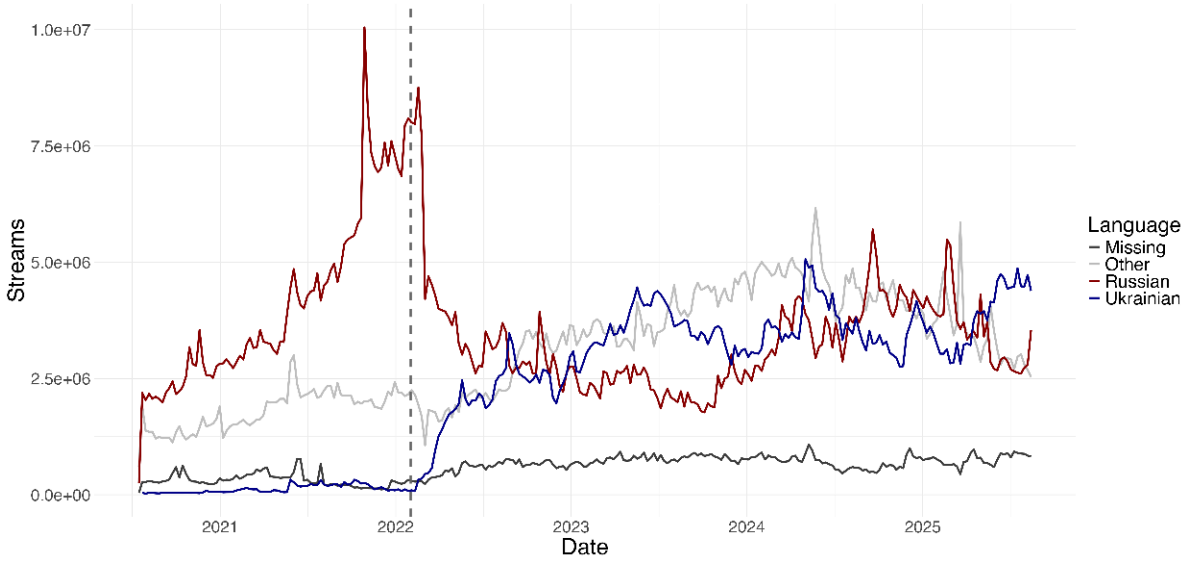
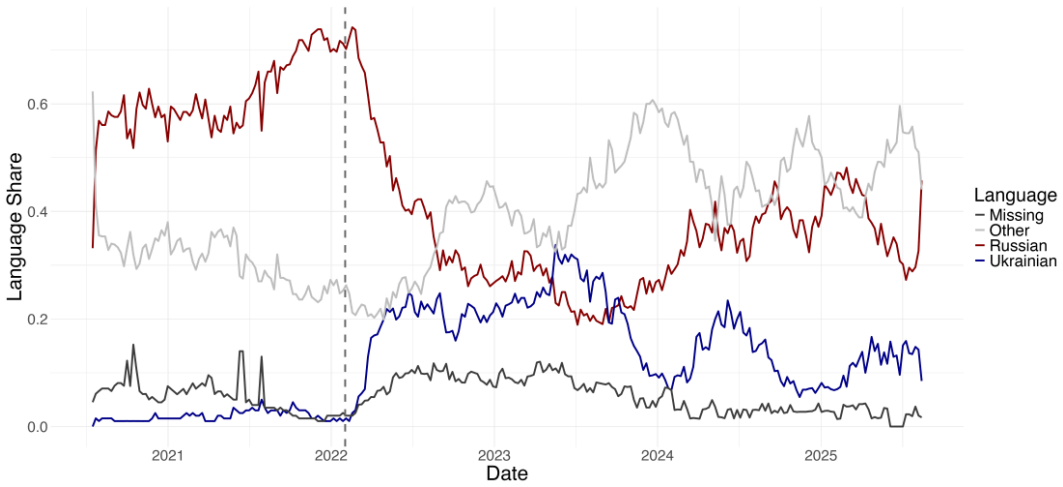
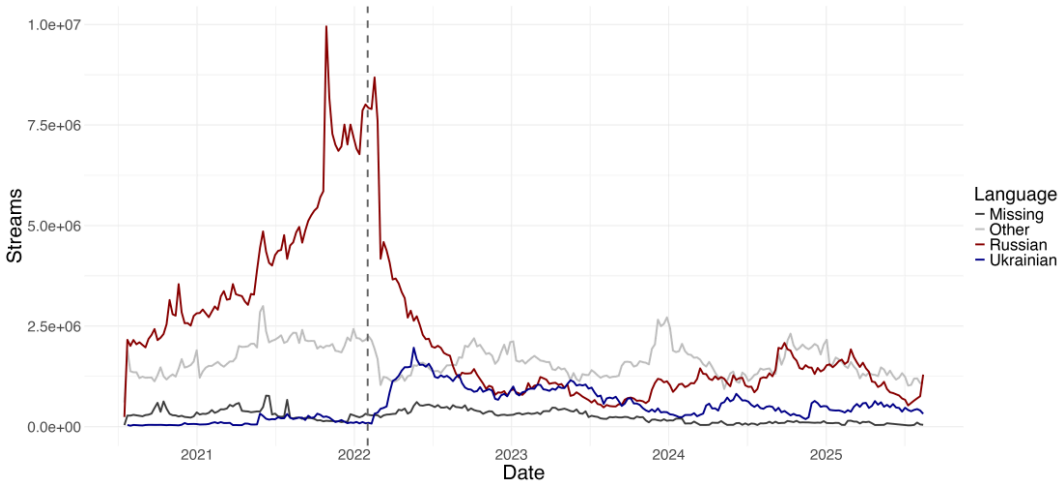


Figure D.11: Total Weekly Streams by Language in Spotify's Top-Songs Chart in Ukraine

Notes: This figure shows how streaming volume in Spotify's weekly chart of the top-200 songs in Ukraine evolved by language of the track (Ukrainian, Russian, or other) before and after the 2022 invasion. It reports total weekly streams for charting tracks from July 2020, when Spotify launched in Ukraine, through August 2025. Track language is determined from lyrics retrieved from Genius; tracks without available lyrics, mostly instrumentals and songs not documented on Genius, are shown separately. The vertical dashed line marks Russia's full-scale invasion in February 2022.



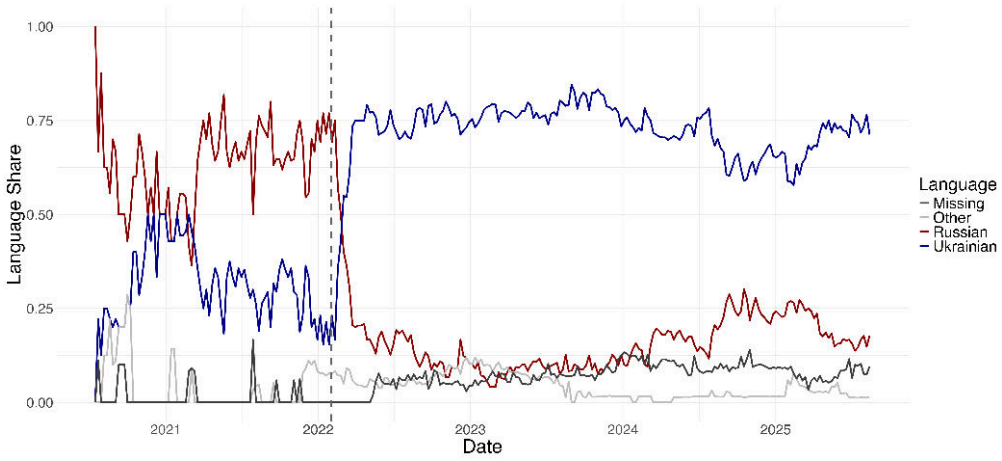
(a) Language Composition of Chart Entries



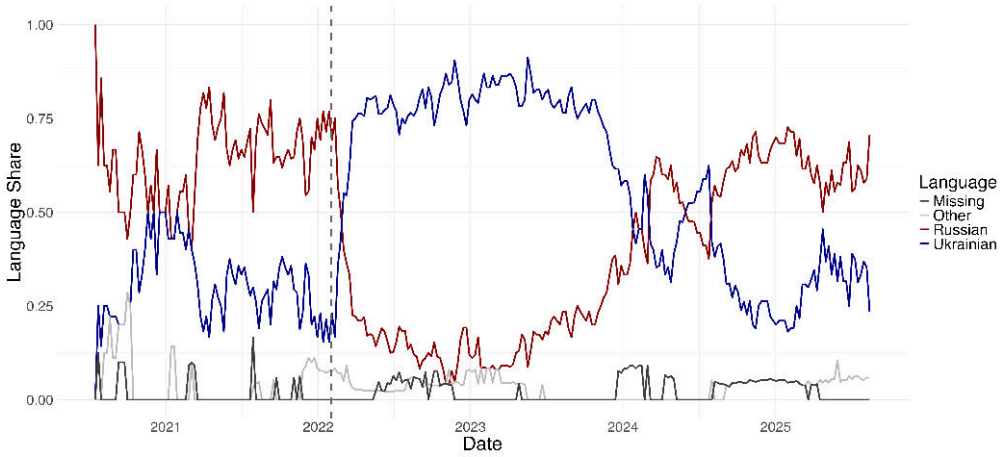
(b) Total Weekly Streams by Language

Figure D.12: Spotify's Top Songs in Ukraine: Tracks Released Before February 24, 2022

Notes: This figure presents analogs of Figure 4 and Figure D.11, restricting the sample to tracks released before February 24, 2022. Panel A shows how the language composition of Spotify's weekly chart of top-200 songs in Ukraine evolved across Ukrainian-, Russian-, and other-language tracks; Panel B shows the corresponding total weekly streaming volume by language. The chart data run from July 2020, when Spotify launched in Ukraine, through August 2025. Track language is determined from lyrics retrieved from Genius; tracks without available lyrics, mostly instrumentals and songs not documented on Genius, are shown separately. The vertical dashed line marks Russia's full-scale invasion in February 2022.



(a) All Releases



(b) Preinvasion Releases

Figure D.13: Language Composition of Spotify's Top Songs in Ukraine: Ukrainian Artists

Notes: This figure presents analogs of Figure 4 for songs performed by artists of Ukrainian origin. Panel A shows how the language composition of Spotify's weekly chart of top-200 songs in Ukraine evolved across Ukrainian-, Russian-, and other-language tracks; Panel B shows the same outcome restricting the sample to tracks released before February 24, 2022. The chart data run from July 2020, when Spotify launched in Ukraine, through August 2025. Track language is determined from lyrics retrieved from Genius; tracks without available lyrics, mostly instrumentals and songs not documented on Genius, are shown separately. The vertical dashed line marks Russia's full-scale invasion in February 2022.

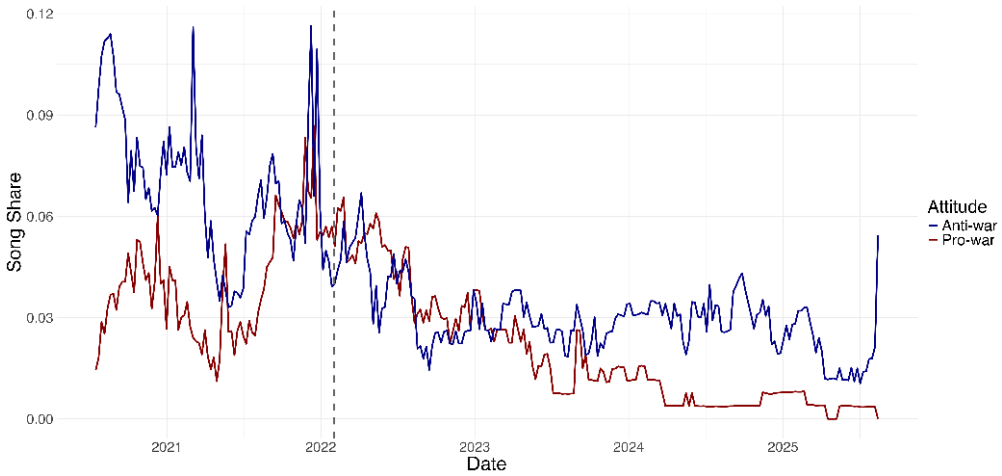
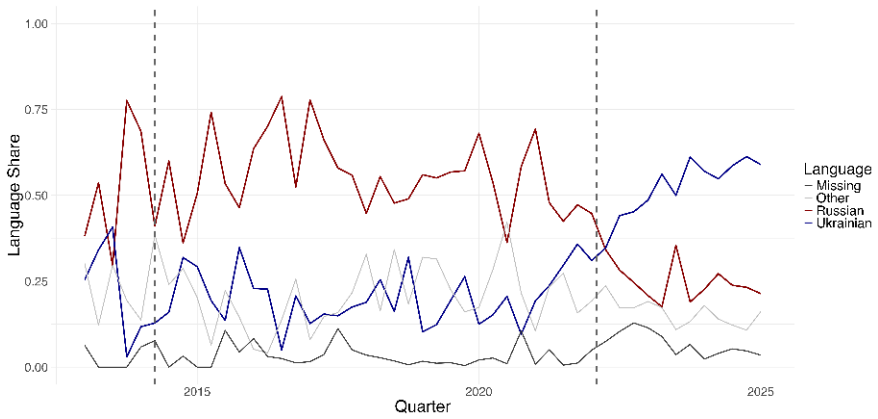
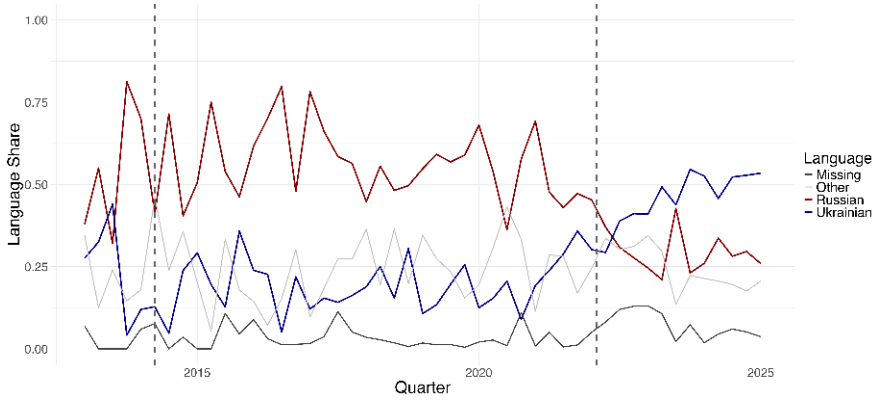


Figure D.14: Spotify Chart Share of Russian Artists by Their Stance on the War

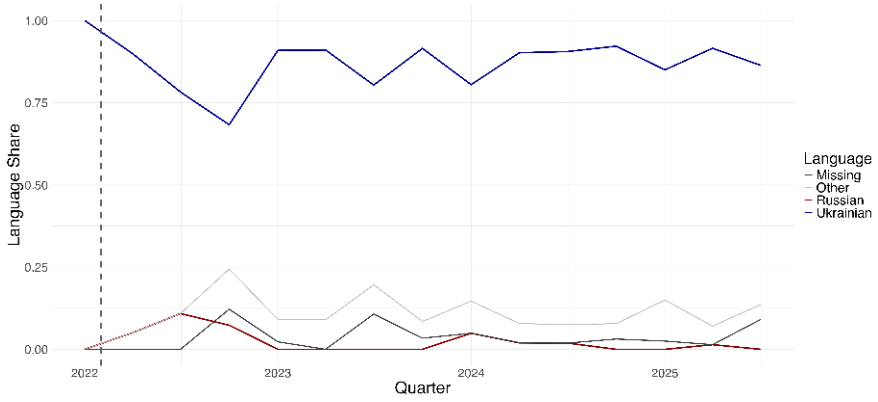
Notes: This figure plots the weekly share of entries in Spotify's weekly chart of top-200 songs in Ukraine accounted for by Russian artists who publicly support the war (prowar) and by those who publicly oppose it (antiwar), as classified by the Stars About War Project (2025). The chart data run from July 2020 to August 2025. The vertical dashed line marks Russia's full-scale invasion in February 2022.



(a) All Releases



(b) Artists Active Before and After the Invasion



(c) New Artists

Figure D.15: Language of New Releases by Ukrainian Artists

Notes: This figure shows how the quarterly share of new releases by artists of Ukrainian origin evolved across Ukrainian-, Russian-, and other-language tracks from 2013 to 2025. Panel A includes all artists; Panel B restricts the sample to artists with releases both before and after February 24, 2022; Panel C restricts the sample to artists whose first observed release occurs after February 24, 2022. Artists of Ukrainian origin are identified from NUAM Club, Spotify, and Wikipedia, and their release histories are retrieved from Genius (Appendix C.4). Track language is determined from lyrics retrieved from Genius; tracks without available lyrics, mostly instrumentals and songs not documented on Genius, are shown separately. Vertical dashed lines mark the onset of hostilities in 2014 and Russia's full-scale invasion in 2022.

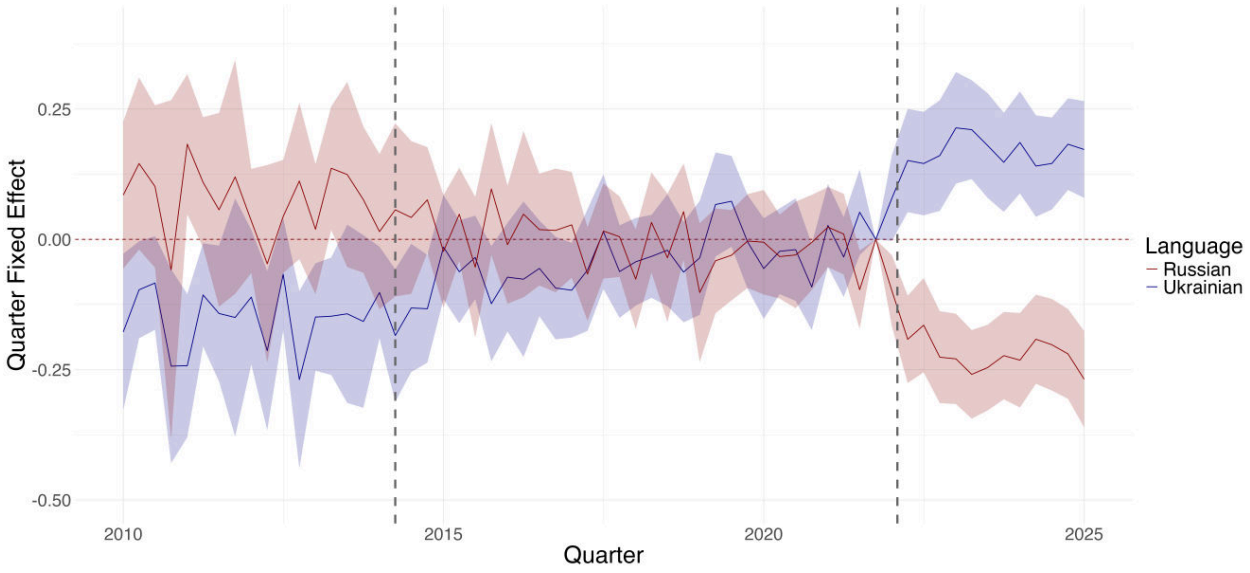
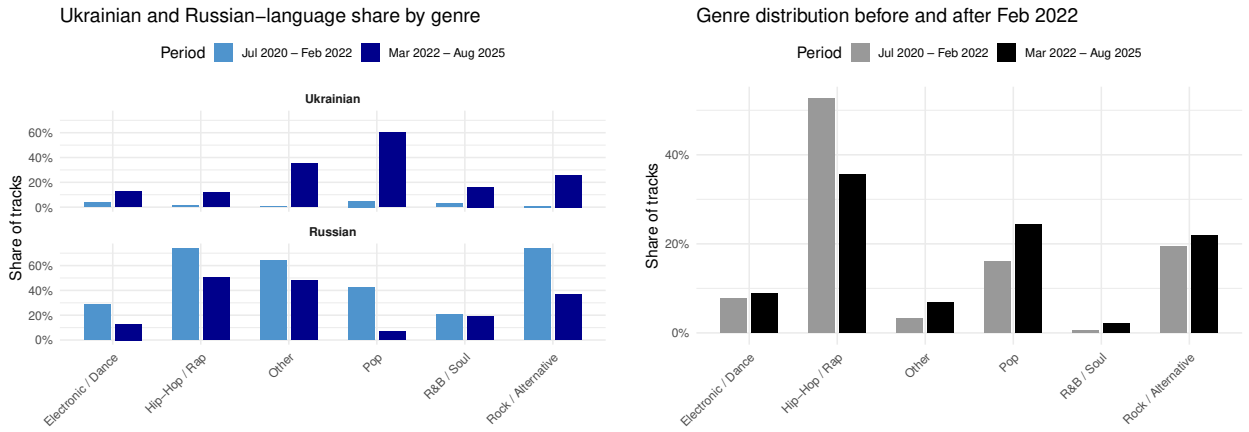


Figure D.16: Within-Artist Changes in Release Language

Notes: This figure plots quarter fixed effects from regressions of the share of an artist's releases in Ukrainian and, separately, in Russian on quarter indicators and artist fixed effects: for each language l, $share^l_{at} = \theta^l_t + \mu_a + \varepsilon^l_{at}$, where $share^l_{at}$ is the share of releases by artist a in quarter t in language l, θ^l_t are language-specific quarter fixed effects, and μ_a are artist fixed effects. The sample consists of artists of Ukrainian origin with releases from 2010 through 2025 (Appendix C.4). Coefficients are normalized to 2021Q4, the last quarter before Russia's full-scale invasion. Vertical dashed lines mark the onset of hostilities in 2014 and Russia's full-scale invasion in 2022. Standard errors are clustered at the artist level; shaded areas denote 95% confidence intervals.

Genre composition. Theoretically, the observed shift toward Ukrainian songs could be driven not by an actual increase in popularity but rather by the public visibility of certain musical genres—for instance, the result could stem from music that plays primarily in public settings, such as electronic/dance tracks. To rule out this confounding factor, we scraped genre metadata for all artists in our dataset via the Last.fm (2024) API and classified them into six broad genres: Hip-Hop/Rap, Pop, Electronic/Dance, Rock/Alternative, R&B/Soul, and Other. We then examine the change in the share of Ukrainian songs within these genres, as well as the overall popularity of the genres themselves, comparing the preinvasion period (July 2020–February 2022) to the postinvasion period (March 2022–August 2025). The results are presented in Figure D.17. The popularity of Ukrainian songs experiences a sharp increase across all categories, jumping from less than 5% before the full-scale invasion to between 10% and 60%, depending on the genre. Meanwhile, the share of Russian songs drops in all genres, decreasing by more than half in Pop, Electronic/Dance, and Rock/Alternative. Overall genre composition remains almost unchanged throughout the sample period, except for a shift in listening shares from Hip-Hop/Rap to Pop.



(a) Language Shares Within Genre (b) Genre Popularity Shares

Figure D.17: Genre Popularity in Ukraine, Before and After the Invasion

Notes: This figure examines how the language composition of charting songs varies across musical genres before and after Russia's full-scale invasion, and whether the overall genre composition of Spotify's weekly chart of top-200 songs in Ukraine changed over the same period. The preinvasion period is July 2020–February 2022 and the postinvasion period is March 2022–August 2025. Artists are assigned to six broad genres (Hip-Hop/Rap, Pop, Electronic/Dance, Rock/Alternative, R&B/Soul, and Other) using genre metadata from the Last.fm API; track language is determined from lyrics retrieved from Genius. Panel A reports, within each genre, the share of charting tracks in Ukrainian and Russian in each period. Panel B reports the share of chart entries accounted for by each genre in each period.

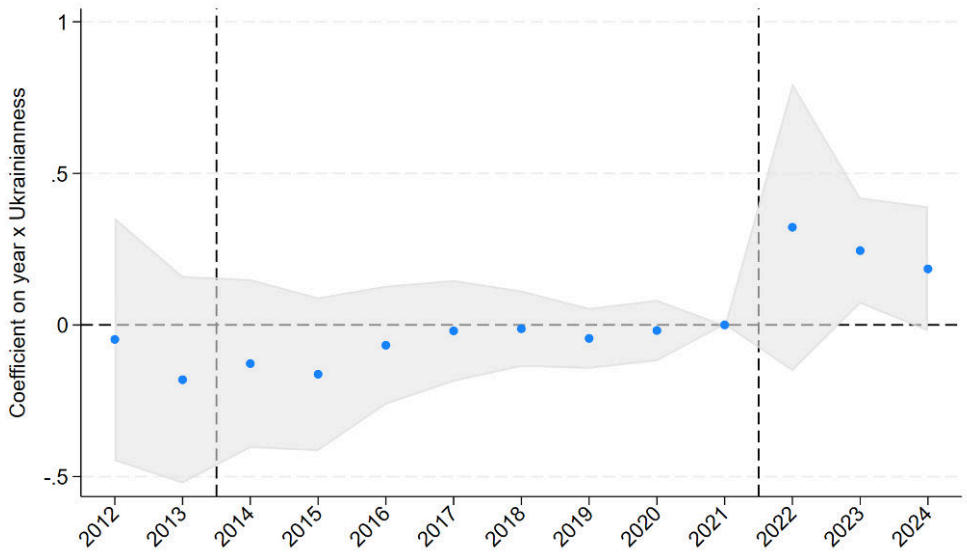
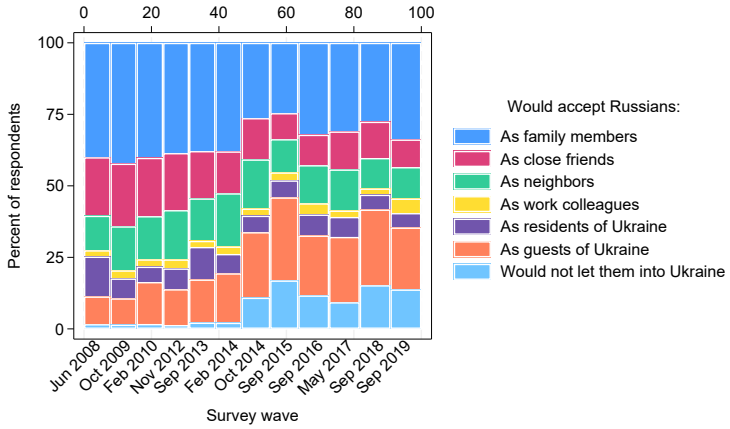
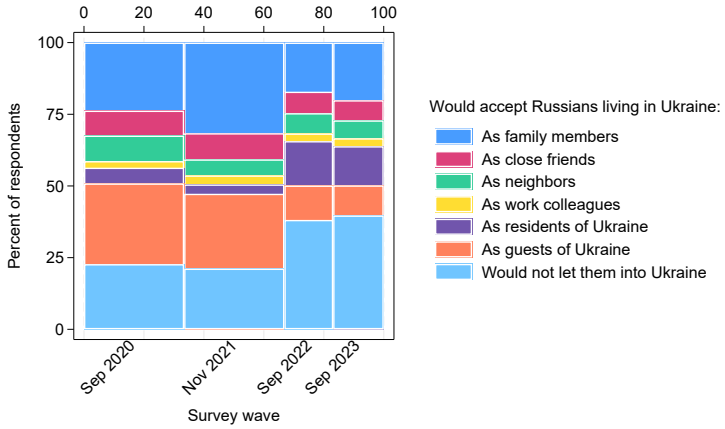


Figure D.18: Cuisine Searches: Ukrainianness and Standardized Dish Popularity

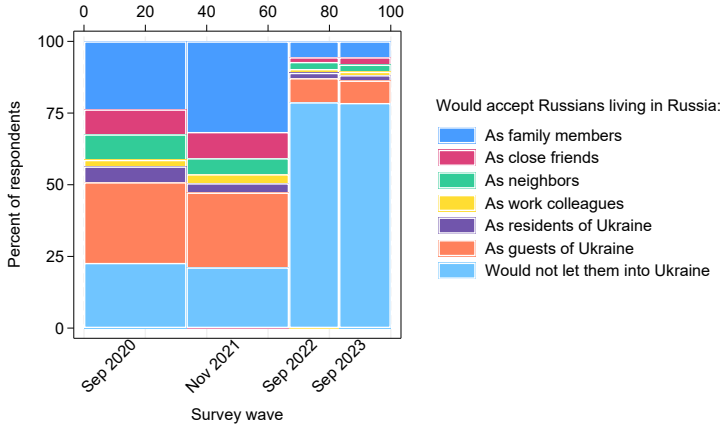
Notes: This figure re-estimates Panel B of Figure 5 in the main text after standardizing the Google Trends score of each dish by its own prewar mean and standard deviation. Coefficients are on interactions between year indicators and the dish's prewar Ukrainianness score, normalized to 2021, the omitted reference year; the specification includes dish-by-oblast and oblast-by-month fixed effects. Vertical dashed lines mark the onset of hostilities in 2014 and Russia's full-scale invasion in 2022. Standard errors are two-way clustered by dish and oblast; shaded areas denote 95% confidence intervals.



(a) Russians, 2008–2019



(b) Russians Living in Ukraine, 2020–2023



(c) Russians Living in Russia, 2020–2023

Figure D.19: Attitudes Toward Russians

Notes: The figure reports the distribution of responses to questions about whether respondents would accept Russians in various social roles, from accepting them as family members to not allowing them into Ukraine. Panel A uses a single "Russians" measure between 2008 and 2019. Beginning with the 2020 questionnaire, the question distinguishes between Russians living in Ukraine (Panel B) and Russians living in Russia (Panel C). The two measures coincide in the preinvasion waves and diverge afterward.

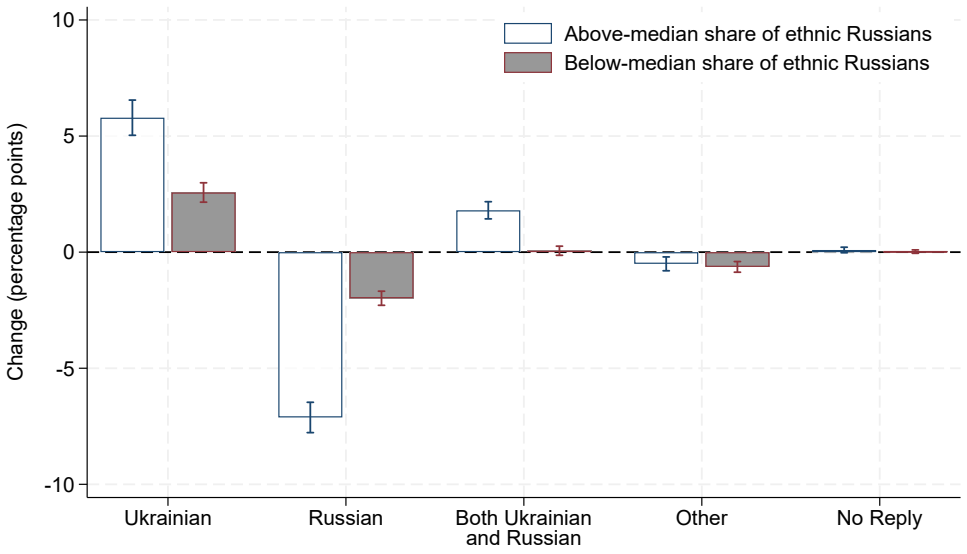


Figure D.20: Change in Reported Ethnicity Post 2014 War, by Prewar Ethnic Shares

Notes: This figure plots the change in the mean percentage of KIIS respondents naming Ukrainian, Russian, both Ukrainian and Russian, or another ethnicity in the open-ended nationality question, or not replying, between the period after the 2014 war (March 2014–February 2020) and the period before it (June 2008–February 2014): $\Delta\bar{x} = \bar{x}_{\text{Mar 2014–Feb 2020}} - \bar{x}_{\text{Jun 2008–Feb 2014}}$. White bars are respondents in rayons with an above-median prewar share of ethnic Russians, measured using the 2001 Census; gray bars are respondents in rayons with a below-median share. The sample excludes Crimea and rayons fully or partially occupied after 2014. Error bars denote 95% confidence intervals.

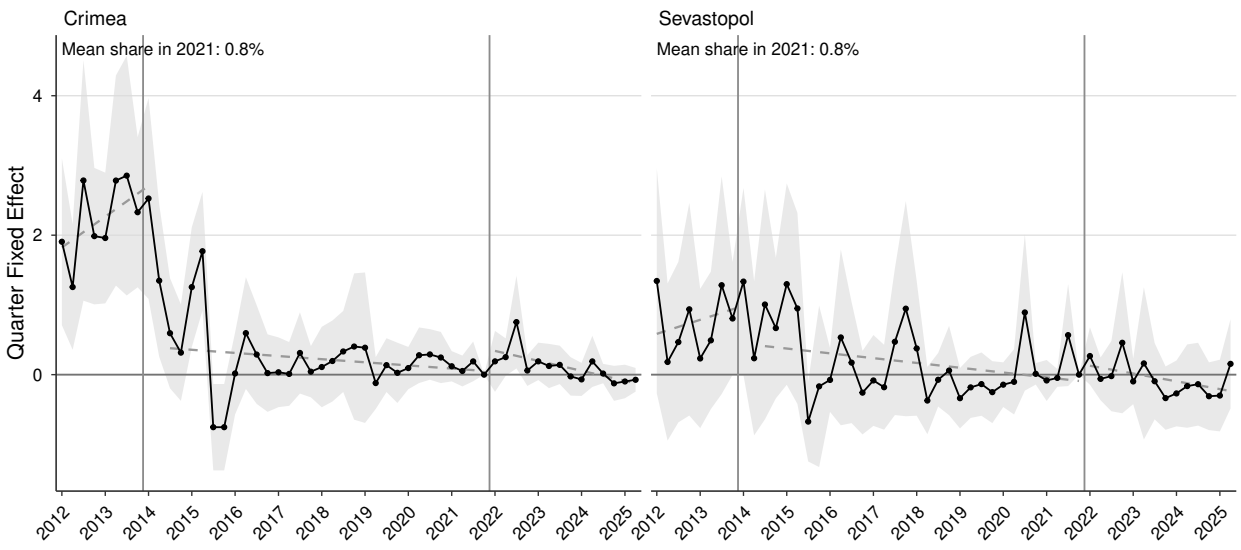


Figure D.21: Ukrainian-Language Share of Google Searches in Crimea and Sevastopol

Notes: This figure plots quarterly time fixed effects from regressions of the share of searches using the Ukrainian spelling within each Ukrainian–Russian word pair on quarter indicators, estimated separately for Crimea (left panel) and Sevastopol (right panel) from 2012Q1 through 2025Q2, in percentage points. The sample includes 198 word pairs. Coefficients are normalized to 2021Q4, the last quarter before Russia's full-scale invasion. Vertical lines mark the onset of hostilities in 2014 and Russia's full-scale invasion in 2022, drawn between the adjacent quarters; dashed sloped lines are linear trends fitted within each subperiod. The annotation in each panel reports the mean Ukrainian-spelling share of searches in 2021. Standard errors are clustered by word pair; shaded areas denote 95% confidence intervals.

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*.

Assouad, L. (2025). Leadership and Nation-Building. Working Paper.

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*(4), 1193–1251.

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 (2026). Religion, Education, and the State. *Review of Economic Studies 93*(3), 1494–1535.

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

Blanc, G. and M. Kubo (2025). The Making of France.

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

Bureau of National Statistics of Kazakhstan (2026). Popular names. https://stat.gov.kz/ru/instuments/name/. Accessed August 2026.

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.

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

Dahis, R., E. Nix, and N. Qian (2019). Choosing Racial Identity In the United States, 1880-1940. Technical report, National Bureau of Economic Research.

de Figueiredo Jr, R. 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.

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.

Dekel, I. (2026). Social identity and dissent. *Working paper*.

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.

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.

Fryer, R. G. and S. D. Levitt (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics 119*(3), 767–805.

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/.

Gidron, N. and M. Shayo (2026). The political economy of national identity. In G. A. Akerlof and R. E. Kranton (Eds.), *Handbook of the Economics of Identity*. Forthcoming.

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).

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.

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

Kikuchi, M. (2024). How does war affect cultural tolerance? evidence from concert programs, 1900–60. *Journal of Peace Research 61*(2), 163–179.

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.

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

Last.fm (2024). Last.fm web services api. https://www.last.fm/api.

Leipzig Corpora Collection (2025). Leipzig corpora collection.

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

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.

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/.

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.

Qian, N. and M. Tabellini (2026). Racial discrimination and the social contract: Evidence from US Army enlistment during WWII. *Review of Economic Studies 93*(2), 1261–1295.

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. (2020). Social identity and economic policy. *Annual Review of Economics 12*, 355–389.

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.

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

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

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

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

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 (2025). Category: Singers of Ukraine.

Zhukov, Y. M. (2022). VIINA: Violent Incident Information from News Articles on the 2022 Russia–Ukraine War. GitHub repository.