Censorship in Democracy

Marcel Caesmann[*], Matteo Grigoletto[†], Lorenz Gschwent[‡]

June 18, 2026

Abstract

Democracies increasingly use censorship to counter foreign propaganda, yet evidence on its consequences remains scarce. We exploit the European Union's 2022 ban on Russia Today (RT) and Sputnik as a natural experiment, using a triple-difference design that compares users connected to the banned outlets against unconnected users in EU and non-EU countries. We analyze a daily panel of 677,780 tweets from 146,633 Twitter users in seven European countries. Pro-Russia output declines by 21.7% among connected EU users; in a difference-in-differences comparison with non-EU users, total pro-Russia output among EU users falls by 13.6%. Alternative suppliers do not fill the gap: neither their pro-Russia output nor the engagement they receive rises after the ban. Consistent with an agenda-setting role of the banned outlets, the share of EU users' tweets covering the outlets' daily top-five topics decreases by 17%. A survey experiment offers suggestive evidence that such censorship can come at a cost to the very norms it is meant to defend.

Keywords: Censorship, Propaganda, Text-as-data, Media slant
JEL Classification: D72, D78, L82, P16

Acknowledgments:
An early version of this paper was co-authored with Janis Goldzycher, we thank him for his superb work. We are grateful to Joop Adema, Elliott Ash, Milena Djourelova, Ruben Durante, Ruben Enikolopov, Kai Gehring, Gaël Le Mens, Ro'ee Levy, Pierre-Guillaume Méon, Carlo Schwarz, Marta Serra-Garcia, Joachim Voth, David Yanagizawa-Drott, Nikita Zakharov, and seminar participants at the University of Zurich, Bocconi University, University of Bern, RTG Regional Disparities and Economic Policy, University of Duisburg, as well as participants and discussants at the CINCH Research Agenda for Ukraine workshop, 5$^{\text{th}}$ Monash-Warwick-Zurich Text-as-Data workshop, Annual Meeting of the Austrian Economic Association (NOeG) 2023, Digital Democracy Workshop DigDemLab Zurich, 17$^{\text{th}}$ CESifo Workshop on Political Economy, 2$^{\text{nd}}$ CESifo/ifo Junior Workshop on Big Data, 7$^{\text{th}}$ Conference on Political Economy of Democracy and Dictatorship, 17$^{\text{th}}$ RGS Doctoral Conference, European Public Choice Society 2024, FRIAS conference Understanding the Rise of Autocrats in the 21$^{\text{st}}$ Century, for their helpful comments. Marcel Caesmann thanks the UBS Center for Economics in Society and the University of Zurich Graduate School Director's Grant for generous support of this project. Matteo Grigoletto thanks the Wyss Academy for Nature at the University of Bern for the support of this project. The authors gratefully acknowledge the computing time granted by the Center for Computational Sciences and Simulation (CCSS) of the Universität of Duisburg-Essen and provided on the supercomputer amplitUDE (DFG project 459398823; grant ID INST 20876/423-1 FUGG) at the Zentrum für Informations- und Mediendienste (ZIM).

[*] University of Zurich, e-mail: *marcel.caesmann@econ.uzh.ch*
[†] University of Bern, Wyss Academy for Nature at the University of Bern, e-mail: *matteo.grigoletto@unibe.ch*
[‡] University of Duisburg-Essen, e-mail: *lorenz.gschwent@uni-due.de*

1 Introduction

Slanted media can be very effective in swaying opinions and behavior (Enikolopov, Petrova, and Zhuravskaya 2011; Yanagizawa-Drott 2014; Adena et al. 2015). Recognizing this, autocratic regimes extensively use media censorship to suppress dissent and manage the flow of information within their borders (Guriev and Treisman 2022). Increasingly, however, these same regimes harness the power of media and redirect it outward, deploying state-controlled outlets, coordinated narrative messaging, and disinformation to exert influence abroad. Democratic societies are vulnerable to this threat – one made more acute as a growing share of the public comes to rely on social media as a primary source of information,[1] with mounting evidence that slanted and false content on these platforms can distort political outcomes, deepen affective polarization, and fuel hate crimes (Lazer et al. 2018; Allcott et al. 2020; Levy 2021; Müller and Schwarz 2023). As the scale of the threat has become clearer, the search for effective policy responses has grown increasingly urgent (Persily and Tucker 2020), pushing democracies toward measures once considered exceptional. Yet these responses carry an inherent trade-off: countering foreign propaganda may require restricting speech, the freedom democracies are built to protect.

Assessing this trade-off requires weighing two factors, neither of which has received systematic empirical attention. The first concerns effectiveness: because democratic societies protect freedom of information, a motivated audience can seek out alternative suppliers of restricted content, so it is far from obvious whether removing a source reduces exposure to its narratives or merely redirects it. The second concerns cost: free speech and media independence are widely regarded as constitutive of liberal democracy,[2] so a democracy that deploys censorship risks eroding the very legitimacy that sets it apart from the regimes whose instruments it borrows – the tension Popper (1945) termed the paradox of tolerance.

Democracies have two broad policy responses. One operates at the individual level – media-literacy campaigns, fact-checking, and behavioral nudges – and has been evaluated extensively and shown potential to reduce the circulation of false news (Pennycook et al. 2020; Pennycook and Rand 2022; Kozyreva et al. 2024); Guriev et al. (2023) provide a unified framework for assessing such small-scale policies. The other relies on top-down legislative action to regulate platform content: measures such as the US Protecting Americans from Foreign Adversary Controlled Applications Act (2024), the EU's Digital Services Act (2023), and the German NetzDG (Jiménez Durán, Müller, and Schwarz 2025) are becoming a standard instrument of democratic governance, as Israel's ban on Al Jazeera further illustrates. Yet while the individual-level arm is well understood and platform-level moderation is increasingly studied, systematic evidence on direct state bans of foreign media outlets in a democracy is lacking.[3] It is this second, state-led

[1] As of September 2025, 5.41 billion, or 65.7 percent of the world population, use social media (Datareportal, 2025). Overall, just over half of U.S. adults (54%) say they at least sometimes get news from social media (Pew Research, 2024).

[2] Since their inclusion in the First Amendment in 1791, freedom of speech and of the press have been viewed as essential to self-government, though their scope has expanded through modern jurisprudence to become central safeguards of democratic deliberation (e.g., New York Times v. Sullivan, 1964). Especially after 1945, political theory and constitutional practice converged in treating freedom of speech and independent media not only as individual rights but as defining institutions of liberal democracy, with Popper (1945) recognizing them as core features of "open societies".

[3] There is some evidence on censorship in authoritarian regimes (Chen and Yang 2019; Becker, Pino, and Vidal-Robert 2021). For platform-level moderation in democracies, Andres and Slivko (2021) and Jiménez Durán, Müller, and Schwarz (2025) show that the German NetzDG reduced the spread of hateful posts online; these operate through intermediary liability rather than direct state bans on outlets.

instrument that we study.

To assess the impact of censorship in a democracy, we study an unprecedented event at the EU level. Following Russia's full scale invasion of Ukraine on 24 February 2022, the EU banned Russia Today (RT) and Sputnik – two Kremlin-backed outlets that had operated across member states with substantial reach, presenting themselves as conventional news sources while disseminating narratives used to justify the invasion.[4] RT and Sputnik were central nodes in the online war conversation, and other producers of pro-Russia content drew on them for narratives and cues; their removal, therefore, was a supply shock to this market, not only to the outlets' direct audience. Enacted on 2 March 2022, the ban applied to all channels of communication, from television and radio to online platforms, sharply curtailing the outlets' visibility across the EU virtually overnight.

This setting lets us address both questions empirically. On effectiveness: did the ban reduce pro-Russia content, and did the supply shock propagate – leading other producers to cut their own output in turn? On cost: did deploying censorship in a democracy come at a cost to how citizens perceive their institutions' commitment to free speech and media independence?

We begin with effectiveness, on Twitter (now X), a platform that stood at the center of the information war that erupted alongside Russia's military offensive. Twitter was a primary channel through which RT and Sputnik disseminated their narratives before the ban. In the days following the invasion, it became the main arena for the real-time clash between Russian and Ukrainian framings of the conflict: governments, military units, and civilian witnesses used the platform simultaneously to provide operational updates, counter-narratives, and documentation of events on the ground. Twitter was then – and remains – a pivotal platform for shaping public opinion and fueling political activity (Allcott and Gentzkow 2017; Acemoglu, Hassan, and Tahoun 2018): narratives amplified by radical or influential users can escalate (Müller and Schwarz 2023) and frequently migrate into mainstream and traditional media (Cage, Herve, and Mazoyer 2022).

To measure pro-Russia content on Twitter, we classify each tweet with a large language model (LLM) as pro-Russia, neutral, or anti-Russia, and extract up to five topics via open-ended LLM-based topic tagging. The analysis focuses on tweets labeled pro-Russia and their associated topics. We validate the LLM classification using an alternative measure in the spirit of Gennaro and Ash (2023) and Gentzkow and Shapiro (2011): we embed tweets from Russian and Ukrainian government accounts as reference poles and compute relative cosine similarity between them and each tweet in our sample. The two measures agree closely, and all core results are unchanged.

To estimate the ban's effect, we study the production and dissemination of pro-Russia content by users active in the war conversation, exploiting the natural experiment in two complementary ways. We begin with a triple-difference design that combines variation in pre-ban network proximity to RT and Sputnik with EU/non-EU and pre/post variation. The connected indicator captures users who interacted with the outlets directly or with users who did, a proxy for exposure to their content flow rather than direct viewership. Connection is not randomly assigned; the most connected users are plausibly also the most

[4] The ban was upheld by the General Court of the European Union (*RT France v Council*, T-125/22, 27 July 2022), which assessed it as a proportionate restriction under Article 11 of the EU Charter of Fundamental Rights; the prohibition of "propaganda for war" in international law (ICCPR, UN 1966) featured as supporting reasoning.

ideologically committed. We therefore allow pre-ban pro-Russia leaning and exposure to Russia trade to have their own differential post-ban effects (pre-ban pro-Russia attitudes and local trade exposure, each interacted with the ban), and the decline remains concentrated among users network-closest to RT and Sputnik – roughly −22% for close users against a marginal −8% for distant ones (Figure E.5). Because the specification already absorbs differential responses along these dimensions, the gradient is difficult to attribute to selection on observed pre-ban attitudes.

We prioritize this design for two reasons. First, it provides a first-stage check on the supply-shock interpretation: if there is no effect among the users most exposed to the banned outlets, a broader decline in pro-Russia content among the general population would be hard to attribute to the ban. Second, it offers the most credible identification: the triple interaction isolates changes specific to connected EU users, absorbing any shift common to all EU users or all connected users. Connected EU users reduce their expected pro-Russia output by 21.7% after the ban. Event-study pre-trends are flat and statistically indistinguishable from zero, supporting the parallel-trends assumption.

To translate the connected-user response into a population-wide magnitude, we estimate a two-way difference-in-differences comparing all EU to all non-EU users: expected pro-Russia output among EU users falls by 13.6%, while total tweet volume falls only modestly (by 6.7%), indicating that the decline reflects a shift in what users post rather than a withdrawal from the war-related conversation.

The aggregate decline is driven almost entirely by residual posting rather than by direct outlet citation: direct citations of RT and Sputnik fall by roughly 68% but come from a handful of users who are a negligible share of pro-Russia output population-wide, while residual pro-Russia content – posts that never name the outlets – falls by 13.1% across the full panel. Nor do alternative producers fill the gap: pre-ban secondary suppliers – users who posted pro-Russia content on at least four of the eight days before the ban – reduce their own pro-Russia output by 14.5% rather than ramping it up, and the per-tweet engagement they receive does not rise after the ban.

The two-way comparison rests on a stronger assumption than the triple-difference: EU-specific shocks around the invasion – refugee inflows, the broader shift in European sentiment toward Russia – could depress pro-Russia posting among all EU users on their own. The proximity gradient weighs against this reading. A shock common to all EU users, regardless of network position, would not produce a decline concentrated near the banned outlets and robust to conditioning on pre-ban attitudes and trade exposure (Section 7, Figure E.5).

Turning to the mechanism, we provide evidence consistent with RT and Sputnik having acted as agenda-setters for the broader pro-Russia ecosystem. We argue this role operated through two channels: the outlets supplied other producers with ready-made narratives that lowered the cost of producing pro-Russia content, and they offered a state-backed seal of approval marking narratives as consistent with the Kremlin line. If the outlets set the agenda, then before the ban, EU users producing pro-Russia content could have tracked their daily editorial choices in real time – which stories to emphasize, which frames to push, which themes to lead with – and once access and visibility were restricted, that tracking would have become much harder. We test this by examining users' topical diversity (the number of distinct topics covered per tweet, both across all content and within pro-Russia content specifically) and their topical alignment with RT and Sputnik's daily top-five topics, computed dynamically day by day from the outlets' own output. Topical diversity is essentially unchanged post-ban: users keep posting about a similarly wide range of subjects. But the share of EU users' tweets covering RT and Sputnik's daily top-five topics falls by 17% relative to non-EU users. We read this as consistent with an agenda-setting mechanism: the removal of a key agenda-setter increased the search cost for others to produce pro-Russia narratives.

We then turn to the second question, the cost of censorship to the norms it is meant to protect (Popper 1945; Linz 1978). We conduct a complementary online survey experiment with 900 participants in three EU countries (Germany, France, and Italy), randomly assigning half to receive information about the 2022 ban on RT and Sputnik. Treated respondents report lower perceived EU commitment to freedom of speech and media independence (pooled into an index) by about 0.09 of a standard deviation (significant at the 10% level). In exploratory (non-pre-registered) heterogeneity analysis, the response is concentrated among self-declared centrists, while respondents on the left or right show no comparable reaction. These results suggest that censorship in a democratic context can come at a cost, registered in how citizens perceive their institutions' commitment to free expression.

Our analysis has limitations on both the observational and the experimental side. The post-ban observation window is short, two weeks, bounded both by Twitter API restrictions for academic researchers[5] and by the UK's own ban of RT and Sputnik on 18 March 2022, which would compromise our control group beyond that date. The design, therefore, speaks to the immediate response to the ban rather than its longer-run trajectory. The corpus is restricted to war-related tweets on a single platform, leaving open how the ban affected other dimensions of pro-Russia discourse and whether content migrated to channels we do not observe. The complementary survey experiment was fielded in 2025 and therefore measures a salience treatment on perceptions that have had time to settle, rather than the immediate response to the ban in 2022. To our knowledge, ours is nonetheless the first quasi-experimental study of a direct state ban on foreign media in a democracy to provide evidence on both sides of the censorship question – whether it works and what it is perceived to cost.

Related Literature: Our paper contributes to several strands of the literature. First, we contribute to the literature on censorship. Although systematic evidence remains limited, scholarly and policy attention to censorship has grown in recent years, including in democracies, where debates over information control have become more salient. Conceptual works by Shadmehr and Bernhardt (2015) and Gehlbach and Sonin (2014) highlight the trade-off autocratic rulers face between suppressing information and allowing unbiased reporting: censorship can itself signal an attempt at control. News consumers may distance themselves from outlets that do not meet their informational needs. Empirical evidence on the effects of censorship remains scarce. In autocratic contexts, Chen and Yang (2019) study the impact of providing citizens with access to uncensored internet in an experimental setting. While they find that greater freedom to access uncensored content shifts beliefs, attitudes, and intended behavior, these effects arise only when nudges encourage participants to consume otherwise censored outlets. Becker, Pino, and Vidal-

[5] Twitter/X removed free API access for academic researchers in February 2023, during the course of this study, eliminating the possibility of extending our data collection.

Robert (2021) and Blasutto and de la Croix (2023) find that censorship imposed by the Catholic Church during the Counter-Reformation effectively limited the diffusion of Protestant content; however, it also hindered the diffusion of knowledge and induced a reallocation of talent toward compliant activities, thus impeding growth. In Russia, Simonov and Rao (2022) show that outlet-specific characteristics attract readers to government-controlled media and that, once there, readers rarely switch sources. In democratic settings, Bjørnskov and Voigt (2021) examine constitutional provisions preventing media censorship after terror attacks, and Kellam and Stein (2016) show that strong presidents can threaten media freedom even in democracies.

A growing strand of this literature examines subtler, state- or platform-led rules that shape flows of information. Guriev and Treisman (2022) term this selective information withholding "selective censorship". Empirical analyses of this form of censorship have been rare to date. Corduneanu-Huci and Hamilton (2022) find in a cross-country analysis that media outlets that likely reach the median voter have a higher chance of being censored in both autocracies and democracies. Our paper provides the first quasi-experimental study, to our knowledge, of a direct state ban on foreign media in a democracy to address both its effectiveness and its perceived cost. The setting allows us to draw conclusions about a short-term, high-stakes institutional reaction, unlike past work that has largely focused on long-term effects of censorship (Becker, Pino, and Vidal-Robert 2021). We trace adjustments of the supply and observed engagement for slanted content when the dominant suppliers are removed – including the absence of substitution by alternative pro-Russia producers and evidence consistent with an agenda-setting role behind the residual posting response. We complement this with survey-experimental evidence on the perceived cost of censorship in a democratic context.

Second, we also add to the rich literature investigating the political economy of social media (see Campante, Durante, and Tesei (2023) and Aridor et al. (2024) for overviews). Some scholars argue that offline segregation exceeds online segregation (Gentzkow and Shapiro 2011) and that social media can reduce polarization (Barbera 2014). However, much of the evidence documents adverse effects of social media's rise: it is linked to the spread of populism (Campante, Durante, and Sobbrio 2018; Guriev, Melnikov, and Zhuravskaya 2021) and xenophobia (Bursztyn et al. 2019), increasing political polarization (Halberstam and Knight 2016; Levy 2021; Müller and Schwarz 2023), and reducing subjective well-being (Allcott et al. 2020). Given these negative effects, an emerging literature in this space is concerned with online content moderation (Jiménez-Durán 2023) and interventions to counter the spread of false information online (Guriev et al. 2023).

Morales (2020) studies the effect of banning bots programmed to retweet the Venezuelan president Nicolás Maduro's tweets, showing that this makes the discussion on Twitter more critical of the president. Closest to our paper is the work by Müller and Schwarz (2022), who study the effect of banning Trump's account on reducing toxicity among his followers, and by Jiménez Durán, Müller, and Schwarz (2025), who analyze a German regulation requiring platforms to remove online hate speech. In line with our results, both studies show that platform-level content moderation can curb toxic and hateful online speech. Our setting differs from these studies in policy modality rather than in scale alone. Jiménez Durán, Müller, and Schwarz (2025) examines intermediary liability requiring platforms to remove specified content; Müller and Schwarz (2022) examines a single-account deplatforming decision taken by a platform. We study direct state action banning foreign media outlets – a distinct policy instrument with growing salience as democracies confront similar choices about Russian state media, the U.S. legislation on TikTok, and Israel's ban on Al Jazeera. The cross-country dimension of the EU ban also provides a natural control group, complementing existing content-moderation work that has primarily exploited within-country variation in exposure to the regulated content.

Third, our study contributes to the vast economics literature on media slant (Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006) and the effects of propaganda (Enikolopov, Petrova, and Zhuravskaya 2011; Yanagizawa-Drott 2014; Adena et al. 2015), which traditionally examines the impact of increasing propaganda exposure. Our setting allows us to explore the consequences of reducing exposure to media slant, providing a new dimension to the policy debate on media regulation. It also lets us study adjustments in the observed war-related Twitter corpus, both in the supply of slanted content and in audience engagement with that content. Existing work studying changes in the supply of slanted media induced by market entry and changes in political control has established demand reactions of consumers (Durante and Knight 2012; Durante, Pinotti, and Tesei 2019), in line with the belief confirmation motive modeled by the theoretical literature. Our results expand on the existing empirical work on media slant by showing that the removal of prominent suppliers of slanted (pro-Russia) content can also reduce production among other suppliers, consistent with disrupted access to narrative cues and agenda-setting.

The remainder of the paper is structured as follows. Section 2 provides background on the ban; Section 3 presents our conceptual framework. Section 4 describes the data and previews the empirical strategy. Section 5 reports our main results, beginning with the triple-difference design analysis and proceeding with the wider reduced-form difference-in-differences analysis. We then explore potential mechanisms and channels through which the ban unfolded and through which users and institutions reacted in Section 6. We present robustness and placebo tests for our observational analysis in Section 7. Finally, Section 8 presents experimental evidence on the cost of censorship. Section 9 concludes.

2 Setting: The Ban of Russia Today and Sputnik

2.1 The Onset of the War

On 24 February 2022, the Russian Federation launched a full-scale invasion of Ukraine. Three days earlier, the Kremlin had recognized the self-proclaimed "people's republics" of Luhansk and Donetsk, claiming a duty to protect Russian-speaking minorities.[6] In his televised address announcing the attack, President Vladimir Putin insisted that the confrontation reached beyond Ukraine, calling the West an "Empire of lies" that sought to destroy Russia's traditional values.[7] What Russia envisioned as a short-term military operation evolved into a prolonged war, drawing Europe into an unprecedented geopolitical confrontation.

From the outset, the war was fought not only on the battlefield but also in the digital sphere, with

[6] See Figure Ia for a timeline with the most important dates for our study, and Figure Ib for a map of Europe highlighting the countries of interest in our study.

[7] See the article by George C. Marshall: European Center for Security Studies.

Russian state-backed outlets pushing content designed to justify the invasion and erode European support for Kyiv. The EU acted quickly to counter this strategy. On 1 March 2022, the European Council's Decision[8] provided the legal basis to suspend the broadcasting activities of Russia Today (all language services) and Sputnik throughout the EU. One day later, a Council press release confirmed the immediate ban across television, radio, satellite, cable, internet service providers, and social-media platforms. The Council justified the measure by characterizing both outlets as "essential and instrumental" in supporting Moscow's aggression and as a direct threat to public order and security within the EU.

In this section, we address two questions. First, why did the EU single out these two outlets: were they truly influential in shaping wartime discourse on Twitter? Second, did the ban measurably disrupt EU users' access to RT and Sputnik content?

2.2 Motivation of the Ban and Centrality of the Outlets

> *"Systematic information manipulation and disinformation by the Kremlin is applied as an operational tool in its assault on Ukraine. It is also a significant and direct threat to the Union's public order and security. Today, we are taking an important step against Putin's manipulation operation and turning off the tap for Russian state-controlled media in the EU."*
>
> *Josep Borrell on RT and Sputnik*

Borrell's statement, made in his capacity as High Representative of the Union for Foreign Affairs and Security Policy, reflects the EU's view at the time of the Russian full-scale invasion: propaganda and disinformation were seen as operational weapons in the conflict, posing a direct threat to the Union's security. Acting on this assessment, the EU shut off the communication channels of the most prominent Russian state-controlled media, focusing on Russia Today and Sputnik as the primary targets. Whether singling out these two outlets was justified, however, depends on whether they actually occupied a central position in the war-related information ecosystem. We examine this directly by mapping the network of @-mentions in our sample and measuring each outlet's prominence within it.

Figure Ic compares the prominence of Russia Today and Sputnik (pooled as a single entity) with that of established national news outlets in the war-related Twitter network. The underlying dataset is the same one used throughout our analysis, featuring only tweets discussing the war, and identified through the keyword-filtering procedure described in Section 4. We restrict to the pre-ban period to avoid confounding the outlets' network position with the ban's own effect on user engagement. To construct the figure, we extract every @-mention in the dataset. Each mention forms a directed link from the author to the mentioned account; aggregating these links produces a weighted interaction network. For each node, we calculate closeness centrality (rescaled to range from 0 to 1), which measures how easily other users in the conversation reach that account through chains of @-mentions. We draw the comparison outlets from the countries in our sample: news sources from non-EU countries (United Kingdom and Switzerland) are shown in orange; EU outlets (Austria, France, Germany, Ireland, and Italy) in violet.

The figure shows that, before the ban, Russia Today and Sputnik held a prominent position in the

[8] Decision (CFSP) 2022/351 and Regulation (EU) 2022/350.

war-related conversation network. The comparison outlets are handpicked rather than exhaustive, but selected precisely for their standing as major national news providers, making this exercise a meaningful comparison. Among non-EU sources, RT/Sputnik's centrality was roughly half that of *The Telegraph* and *Sky News*, while ranking above *BBC News* and the *Neue Zürcher Zeitung*. Among EU sources, the two banned outlets ranked comparably to *Der Spiegel* and above *Le Monde*, *Welt*, and *Franceinfo*. Russia Today and Sputnik were not marginal outlets singled out for political effect: they were genuinely embedded in the information network through which war-related content circulated, at levels comparable to some of the most established national media in Europe.

2.3 Enforcement of the Ban

Having established that the banned outlets were prominent actors in the war-related conversation, we now ask whether the ban effectively disrupted their interaction with EU users. A measurable, EU-specific disruption is a necessary precondition for the identification strategy laid out in Section 4.3.

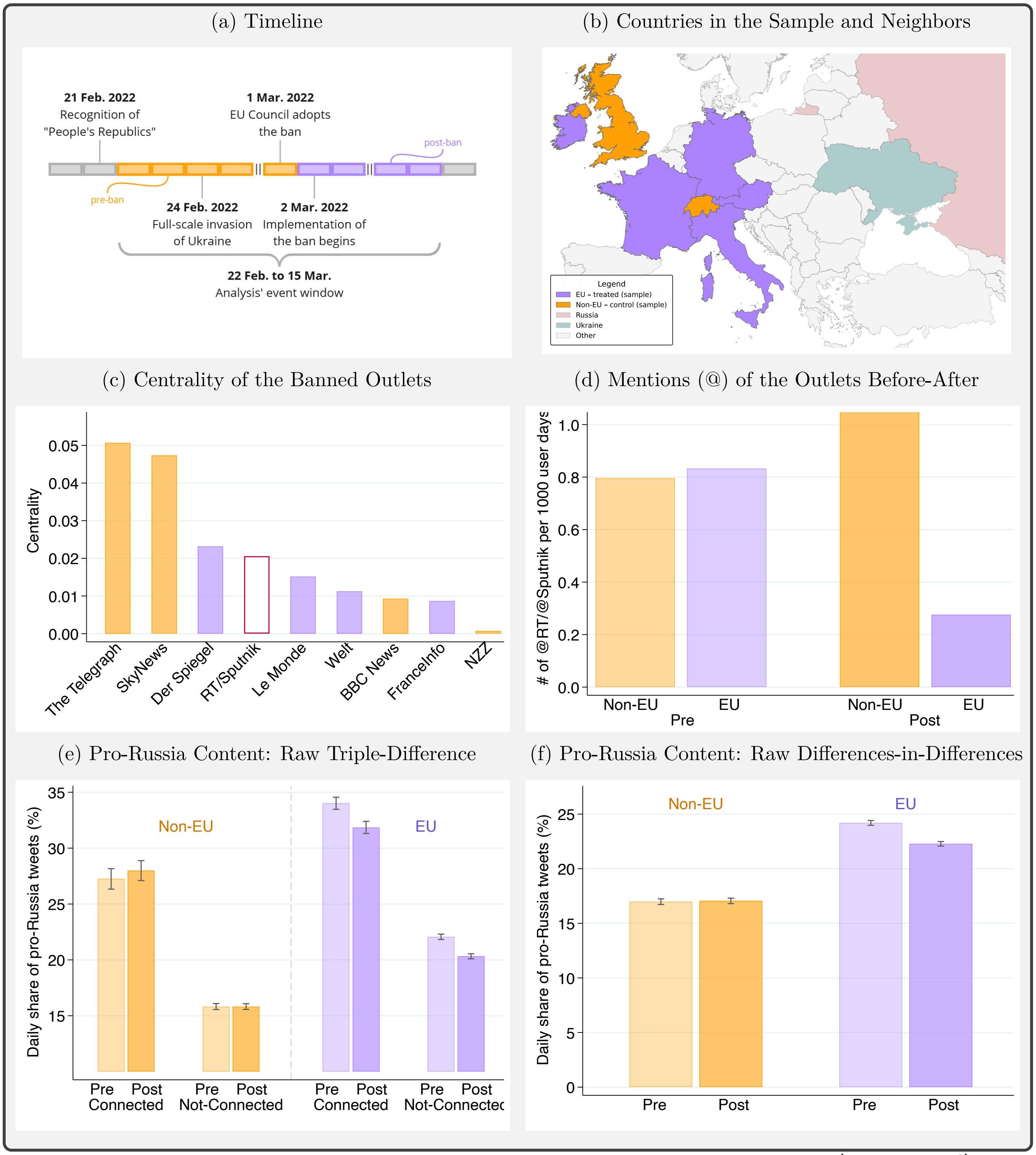
By early 2022, RT and Sputnik already faced restrictions: Meta, Google, and several app stores had demoted or removed their content in some markets – including the United Kingdom – under public and regulatory pressure. The EU measure extended both scope and uniformity: a hard geo-block applied simultaneously across all EU member states, while access in our non-EU comparison countries (the United Kingdom and Switzerland) remained unrestricted on Twitter.

Figure Id tests whether this differential access restriction left a detectable effect on Twitter. Using the same war-related corpus described in Section 4, we count users mentioning the two outlets before and after the ban, separately for non-EU (orange) and EU (violet) users. Mentions are an indirect proxy for accessibility: '@'-handles can be referenced even when the underlying account cannot be retrieved. They remain the best available indicator given Twitter's data constraints.

The contrast is stark. In EU countries, mentions of the banned outlets dropped by more than half once the ban took effect. In non-EU countries, mentions increased by roughly a third over the same window, consistent with broader media dynamics at the time: war discussion intensified after the invasion, and so did engagement with the outlets covering it. Because the EU geo-block did not remove RT and Sputnik's accounts (users could still tag @RT or @Sputnik even without access to their content), the residual EU mentions signal ongoing willingness to refer to the outlets despite the loss of access.

Three caveats deserve explicit acknowledgment here. First, Twitter's enforcement of the geo-block relied on non-public location signals – IP addresses, wireless networks, cell towers – that go beyond the public country setting visible in user profiles and that could not be bypassed by a simple VPN; nevertheless, we cannot rule out that some users may have been successful in circumventing the ban. Our results should, accordingly, be interpreted as intent-to-treat, biased toward zero by any successful circumvention.

FIGURE I.
THE BAN IN CONTEXT: SETTING, RATIONALE, AND DESCRIPTIVE EVIDENCE



Notes: Panel (a) reports the key dates of our study: the analysis window runs from February 22^{nd} to March 15^{th}, 2022. Panel (b) maps the countries in our sample: EU countries in violet, non-EU countries in orange. Panel (c) compares the pre-ban centrality of Russia Today and Sputnik (pooled, red frame) in the Twitter discussion about the war to selected national outlets: non-EU outlets in orange, EU outlets in violet. Panel (d) shows average daily @-mentions of the banned outlets per 1,000 users before and after the ban, separately for non-EU (orange) and EU (violet) users. Panels (e) and (f) show the daily share of pro-Russia tweets before and after the ban, separately for users connected and unconnected to the banned outlets (e) and for all users (f): orange bars indicate non-EU users, violet bars EU users.

Second, while EU users cannot access direct reposts or retweets of RT/Sputnik content (the geo-block restricts retrieval at the EU end regardless of what others share), they remain exposed to content from non-EU users who copy, paraphrase, or relay RT/Sputnik narratives in their own words. This is a genuine indirect channel: if non-EU users actively emulated the banned outlets' output post-ban, EU-connected users could still absorb those narratives without directly accessing the outlets. If operating at scale, this channel would attenuate the ban's impact on the information environment of EU users, even holding their own posting behavior fixed. We address this empirically in Section 6.2: secondary suppliers reduce their pro-Russia output rather than ramping it up to fill the gap, indicating that this indirect channel did not compensate for the outlets' removal within the sampled corpus.

Third, the post-ban surge in non-EU engagement suggests that RT and Sputnik may have reallocated content toward markets where they remained accessible. This does not open an additional exposure channel for EU users (the geo-block holds regardless), but it does affect the comparison group. If non-EU users connected to the outlets received more RT/Sputnik content after the ban and responded by increasing their own pro-Russia output, the non-EU connected group may have received a positive shock. Our estimates should therefore be interpreted as the effect of EU access loss relative to this potentially strengthened non-EU connected baseline, rather than as a no-spillover counterfactual.

The evidence in this section establishes two empirical preconditions for the analysis that follows: the banned outlets were genuinely prominent actors in the war-related discourse, and the ban created a measurable, EU-specific disruption in users' ability to access their content. What is less clear, ex ante, is what this disruption should imply for the production of pro-Russia content by ordinary users and through which channels. In the next section, we lay out a minimal conceptual framework to structure these questions before turning to the data.

3 Conceptual Framework

EU officials described the ban's goal as "turning off the tap" of Kremlin-backed disinformation. The ban directly eliminates RT and Sputnik as active voices on the platform: their own content stops, their material becomes inaccessible, and users can no longer retweet or directly link to their output. But the policy goal was broader: to reduce the production and spread of pro-Russia content among ordinary users, targeting as well the downstream flow of pro-Russia narratives in the broader information environment.

Whether the outlet removal achieves a generalized decrease in pro-Russia content depends on a transmission step that is not automatic. The causal chain of pro-Russia content production runs as follows: the outlets produce content; ordinary users are exposed to it; users amplify, paraphrase, or independently reproduce the narratives they have absorbed; pro-Russia content spreads through the wider conversation. The ban eliminates the first link. Whether the rest of the chain breaks down depends on how reliant users were on the outlets for narrative cues and ready-made content, and on whether removing those cues is enough to disrupt broader pro-Russia production, which in turn hinges on the role RT and Sputnik actually played.

Before the ban, RT and Sputnik functioned as agenda-setters for the broader pro-Russia content ecosystem. While other channels certainly existed, two stand out as the most prominent and directly testable:

(i) **Narrative supply**. The outlets delivered a steady stream of ready-made text, images, and videos that could be paraphrased or translated, introducing talking points and defining which aspects of the war were newsworthy. By lowering the cost of content production, they enabled other suppliers to post pro-Russia content with minimal independent effort.

(ii) **Seal of approval**. Although formally "independent", both brands were widely understood to convey messages consistent with the official Kremlin line, providing users with a low-cost way to verify that a given narrative was state-endorsed.

Together, these two channels allowed RT and Sputnik to establish the topics and framings through which the war was depicted across the pro-Russia content ecosystem. Their removal, therefore, disrupted more than the supply of outlet-generated content: users lost a ready-made source of narratives that shaped what they discussed and how they framed it, and they lost the signal that identified those narratives as state-endorsed.

The net effect of the ban is ex-ante ambiguous, but the characterization above points to two channels through which it may have been effective. Through the narrative supply channel, users face higher costs in finding and framing pro-Russia narratives independently, an effect likely most pronounced among users embedded in the outlets' pre-ban network. Through the seal-of-approval channel, the loss of a clear editorial signal raises uncertainty about what constitutes an approved Kremlin narrative, increasing the cognitive cost of aligned posting. The ban may also have induced self-censorship more broadly, to the extent that users interpreted it as a signal of stricter future moderation — a channel that operates beyond the network of users directly connected to the outlets.

At the same time, several forces may offset these effects. Users may resist the ban, framing the act of posting pro-Russia as an act of protest against censorship. Secondary suppliers such as smaller pro-Russia accounts, foreign outlets, or fringe media may exploit the vacant space left by the banned outlets, compensating for the loss of narrative supply. Finally, the Russian government may adapt institutionally. Diplomats, cultural institutes, bot networks, or non-banned state media could expand their activity, partially replacing the function RT and Sputnik previously served.

These channels are not mutually exclusive; several may operate simultaneously. Which effect prevails is ultimately an empirical question. Our analysis addresses it on three fronts: we document the ban's aggregate effect on pro-Russia content production; we examine whether the disruption reaches secondary suppliers or was instead offset by substitution; and we test the agenda-setting mechanism directly by tracing whether the topics introduced by the banned outlets fade from EU users' content after the ban, the most direct empirical indication of the narrative supply and seal-of-approval channels operating together.

4 Data and Methodology

4.1 Sample of Tweets

We construct our main panel dataset using the Twitter Historical APIv2.[9] The API does not filter on location directly, so we infer the location of users from profile and activity signals. We then construct a balanced user-day panel covering the weeks surrounding the ban, restricting the sample to users actively posting during this window. Our data collection and processing pipeline proceeds in six steps:

(i) **Query definition.** We build the keyword query around the main entities involved in the conflict – Russia, Ukraine, and NATO, leading to the following search string: *russ* OR ukrain* OR nato OR otan.*

(ii) **Initial extraction.** We retrieve all tweets matching the query posted between January 24$^{\text{th}}$ and April 4$^{\text{th}}$, 2022, across three daily time windows (9–12am, 3–6pm, and 8–11pm), yielding 7,865,321 tweets from 1,942,979 users.

(iii) **Geolocation.** We apply the geolocation procedure of Gehring and Grigoletto (2025) to identify users located in our seven target countries – Austria, France, Germany, Ireland, Italy, Switzerland, and the United Kingdom – and restrict to users with at least one tweet in the 22-day analysis window (February 22$^{\text{nd}}$ to March 15$^{\text{th}}$, 2022).[10] This yields 146,633 users.

(iv) **Final download.** For the identified users, we systematically download all their tweets matching the query within the analysis window. We retain only un-truncated tweets – original tweets, replies, and retweets of no more than 140 characters – discarding truncated retweets to avoid systematic measurement error in the classification pipeline. This yields a balanced user-day panel of 3,225,926 user-day observations.

(v) **Classification.** We classify each tweet with GPT-4o-mini as pro-Russia, neutral, or anti-Russia, and extract up to five topics per tweet. We provide more details in the next section; the full prompt and implementation pipeline are described in Appendix A.

(vi) **Outlet data.** Separately, we collect all tweets posted by Russia Today, Sputnik, and their subsidiaries over the same period. These serve as inputs to the agenda-setting analysis in Section 6.6 and to the alternative classification procedure described in Appendix C.

Appendix Table B.3 reports tweet-level descriptive statistics: Panel A covers tweets by Russia Today and Sputnik; Panel B covers our main sample of ordinary users. In Panel A, we show that Russia Today

[9] Data collection coincided with the change in ownership of Twitter in late 2022, which led to the near-immediate discontinuation of research API access. This imposed binding constraints on collection time and total tweet volume, shaping several of the sampling decisions we implement.

[10] We restrict the sample to major Western European countries, excluding Eastern European countries bordering Russia or Ukraine, whose differential exposure to the war could confound the EU vs. non-EU comparison.

is more active on Twitter than Sputnik in our analysis window.[11] About 18% of the outlets' tweets cover the conflict with a pro-Russia slant, and they mention, on average, 2.1 topics per tweet.

Panel B shows that among ordinary users' tweets about the Russo-Ukrainian conflict, the pro-Russia share is comparable to that of the outlets, as is the average number of topics per tweet. At the user level, 6.2% are connected to RT or Sputnik under our definition (either interacting with the outlets directly or with users who did; see Subsection 5.1). Most users in the sample are located in the UK, France, Germany, and Italy, with an overall EU-to-non-EU ratio of roughly 3:2.

4.2 Measuring Pro-Russia Content

We now turn to how we define and measure pro-Russia content. The challenge is to define what counts as "pro-Russia" content and to translate that definition into a reliable, large-scale measure. Traditional Natural Language Processing (NLP) methods – dictionary approaches, topic modeling, or language similarity – are too blunt for this nuanced classification task. We instead use a large language model (LLM) to classify tweet content. Specifically, we use OpenAI's GPT-4o-mini, accessed via the OpenAI API. Prompt design is central: it must specify the setting and the meaning of "pro-Russia" without over-constraining the model's judgment. We arrive at the final prompt through iterative testing. Below, we report its first part:

> *"You are an objective political analyst tasked with analyzing tweets related to the Russia-Ukraine war. Your goal is to classify whether a tweet contains pro-Russia content and, if so, identify key topics in the tweet. Respond strictly in JSON format. Context: The tweets you are analyzing were posted within a one-month window around the February 2022 full-scale invasion of Ukraine by Russia. The war has its roots in a broader geopolitical conflict that began in 2014 with Russia's annexation of Crimea and its involvement in the conflict in Eastern Ukraine. Your analysis should be conducted with this historical and political context in mind. Pro-Russia content refers to messaging that supports, justifies, or aligns with Russian interests in the context of this war."*

The set of possible output labels can affect both LLM predictions and their stability (Zhao et al. 2021; Liu et al. 2023). We therefore instruct the model to classify each tweet into one of three mutually exclusive categories – pro-Russia, neutral, or anti-Russia – in the context of the war; the main analysis focuses on the pro-Russia category, with the others used for robustness checks. Figure B.2 reports the distribution across all three labels over time. Alongside the classification, the model extracts up to five topics per tweet, which we use to test the agenda-setting hypothesis in Section 6.6.

We validate the LLM measure with an embedding-based approach in the spirit of Gennaro and Ash (2023) and Gentzkow and Shapiro (2011). For each tweet, we compute the cosine similarity of its sentence-embedding to two reference poles constructed from tweets posted by Russian and Ukrainian government accounts, respectively. The relative similarity to the Russian pole is a continuous measure of pro-Russia slant that is independent of the LLM prompt design, making the two measures useful complements rather

[11] Russia Today used several accounts, tweeting in different languages, including *@RTenfrancais*, *@de_RT_com*, *@RT_com*, *@RTUKnews*, and *@RT_America*, while Sputnik only used one account, *@SputnikInt*.

than substitutes. Appendix Figure C.1 compares the word-stem frequencies in the two pools and shows that the contrast is substantive: stems like *aggression*, *invasion*, and *occupi* concentrate on the Ukrainian side, while *nato*, *west*, and *nazi* concentrate on the Russian side. Appendix C reports the pole construction in detail and replicates our main results using the embedding-based measure; the two measures broadly agree on which tweets carry pro-Russia content, and the substantive findings are unchanged.

4.3 Identification Strategy

The ban on Russia Today and Sputnik provides an ideal quasi-experimental setting. Three main features make it well suited for causal identification. First, the ban was implemented simultaneously across media platforms in all European countries on 2^{nd} March, providing a sharp and well-identified treatment date and location. Second, EU users did not anticipate such a swift institutional response, and the ban arrived as a shock, ruling out pre-treatment anticipatory adjustment by the users.[12] Third, our comparison countries, the United Kingdom and Switzerland, faced no equivalent restriction during the analysis window: comparable measures followed roughly three weeks later, leaving a clean two-week post-ban period for identification.

We exploit this setting through two complementary designs. First, we estimate a triple-difference specification that isolates the supply-side effect of the ban on users most directly exposed to RT and Sputnik: the connected users. A connected user is any user with a direct or degree-two link to either outlet in the pre-ban network: users who either interacted with the outlets directly or interacted with users who did. This design exploits three sources of variation jointly: pre-ban connection to the outlets, EU vs. non-EU location, and pre vs. post-ban timing. The triple interaction nets out any shock common to all EU users and any shock common to all connected users, isolating the differential change among connected EU users that is most plausibly attributable to the ban. We treat this effect as a necessary first stage for the supply-side interpretation: if the users in the closest orbit of the banned outlets were unaffected, attributing any aggregate decline to a disruption of the outlet network would be hard to defend. Other channels, signaling effects or norm shifts, could still operate, but the supply-side reading would lose its anchor.

Second, we widen the lens to the full European user population using a two-way difference-in-differences design that compares EU and non-EU users before and after the ban. This specification answers a different question: not whether the directly exposed network responded, but how large the ban's overall effect was across the EU user base: a reduced-form estimate of its aggregate impact on pro-Russia discourse on Twitter. The two designs are therefore complementary by construction. The triple-difference identifies a clean supply-side response on a targeted subpopulation; the two-way DiD measures the aggregate effect on pro-Russia output, which the subsequent mechanism analysis (Section 6) then attributes to specific channels.

12 The ban has been viewed by some as a critical departure from the EU's prohibition on obligatory online monitoring under Article 15 of the E-Commerce Directive (see the directive). Russia Today challenged the measure before the European Court of Justice, as did the main Dutch journalists' union, which argued that the ban was the wrong way to address Russia's misinformation campaigns (see Reuters' coverage). The European Court of Justice ultimately upheld the ban on the grounds of the prohibition of "propaganda for war".

Panels (e) and (f) of Figure I provide descriptive support for our strategy. Panel Ie shows the average daily share of pro-Russia content for connected and non-connected users in EU and non-EU countries separately. Three patterns stand out. As expected, connected users show a higher share of pro-Russia content than non-connected users in both country groups. In non-EU countries, the share remains virtually unchanged before and after the ban for both user types. Among EU users, however, the share declines after the ban, and the decline is considerably larger for connected users than for non-connected ones. Panel If confirms this picture through a wider lens: comparing EU and non-EU users overall, the pro-Russia share remains stable in non-EU countries while declining clearly among EU users after the ban.

While entirely descriptive, the patterns above motivate the formal analysis in Section 5. For both designs, triple-difference and difference-in-differences, we employ Poisson Pseudo-Maximum Likelihood (PPML), motivated by non-negative count outcomes with substantial mass at zero, and include user and day fixed effects with standard errors clustered at the user level. We develop the triple-difference specification, identifying assumptions and event-study evidence in Section 5.1; we develop the two-way DiD reduced-form in Section 5.2.

5 Main Results

5.1 Identifying the Supply-Side Effect: Triple-Difference

We begin by examining the effect of the ban on users most directly in the orbit of the banned outlets: those connected to RT or Sputnik in the pre-ban interaction network. We define connection as having a direct or degree-two link to either outlet in the pre-ban retweet, reply, and mention network: users who either interacted with the outlets directly or interacted with users who did. We include degree-two users because they were genuinely embedded in the outlet-centered information flow: through their network neighbors, they were likely exposed to RT and Sputnik content and narrative frames, even without interacting with the outlets themselves. The connection indicator should therefore be read as a proxy for exposure to the outlets' content flow rather than as a measure of direct viewership.

The rationale for starting with this subpopulation is straightforward. The ban is, at its core, a supply-side shock that removed two of the most prominent producers of Kremlin-aligned narratives from the European information space. The users most likely to feel this shock are precisely those who relied on the outlets' output as a source of content to amplify, as an agenda-setting signal, or as a low-cost input for their own pro-Russia posts. If the ban left these directly exposed users unaffected, attributing any broader change in EU-wide pro-Russia content to a disruption of the outlet supply would be hard to defend. We therefore treat this result as a necessary first stage for the supply-side reading of the broader effect documented in Subsection 5.2: it must hold for the wider population response to be attributed to disrupted outlet supply rather than to other channels of the ban, such as signaling or norm shifts.

We estimate the following triple-difference specification (Equation 1):

TABLE I. IMPACT ON USERS CONSUMING CONTENT FROM THE BANNED OUTLETS: TRIPLE-DIFFERENCE SPECIFICATION

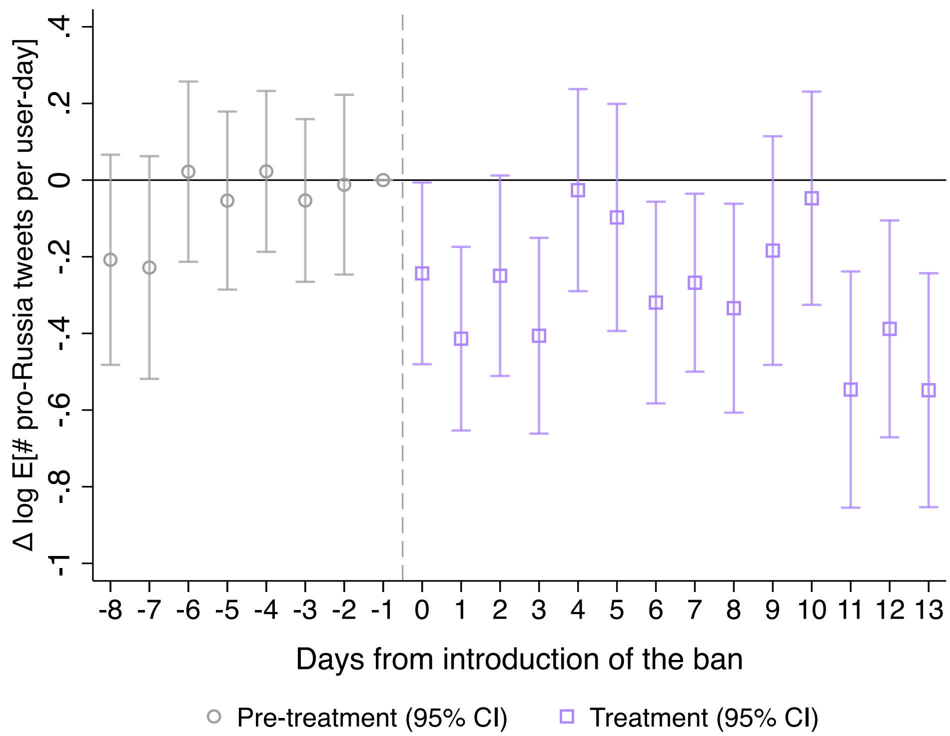
Dependent variable	**No. Pro-Russia Tweets Posted**		
	Coeff./SE/p-value		
	(1)	(2)	(3)
Connected × Ban × EU	-0.148 (0.055) [0.007]	-0.259 (0.073) [0.000]	-0.245 (0.072) [0.001]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Pre-Ban Attitude on Russia: User		✓	✓
Pre-Ban Trade with Russia: State			✓
Pre-Ban Avg. Outcome: Treated	0.498	0.498	0.498
Approx. Percentage Change	-13.77	-22.81	-21.74
Observations	1141514	1141514	1141514

Notes: The table reports results from triple-difference-in-differences models estimated via Poisson Pseudo-Maximum Likelihood (PPML). The outcome variable is the daily count of pro-Russia tweets. The coefficient of interest is the triple interaction between connectivity status, EU membership, and the post-ban indicator – capturing the post-ban change in pro-Russia content production for connected EU users, net of the behavioral change observed among unconnected users in the same period and net of the analogous connected-vs-unconnected contrast in non-EU countries. Connected users are those with direct or degree-two links to RT/Sputnik in the observed pre-ban interaction graph. All columns include user and date fixed effects. Column 2 adds the user's pre-ban share of pro-Russia content, interacted with the post-ban indicator, to allow differential trends by pre-ban attitudes toward Russia. Column 3 additionally controls for the country's pre-ban trade exposure with Russia, also interacted with the post-ban indicator. The sample is the balanced user-day panel restricted to the post-invasion period. Standard errors are clustered at the user level. The reported percentage change is computed as $(e^{\beta}-1)\cdot 100$, where β is the triple interaction coefficient. The pre-ban mean and percentage change rows are computed on each column's own estimation sample.

$$\mathbb{E}[Y_{it} \mid \cdot\,] = \exp\Big[\beta_1\,(Connected_i \times EU_i \times Ban_t) + \beta_2\,(Connected_i \times Ban_t) \\ + \beta_3\,(EU_i \times Ban_t) + \mathbf{X}_i'\boldsymbol{\gamma} \cdot Ban_t + \alpha_i + \delta_t\Big] \quad (1)$$

where Y_{it} is any of our measures of user i's posting behavior on day t: the daily count of pro-Russia tweets, an indicator for any pro-Russia tweet posted, or the count of all tweets about the war. $Connected_i$ is the indicator defined above. EU_i equals 1 for users located in the EU. Ban_t equals 1 on and after 2 March 2022. α_i and δ_t are user and date fixed effects, absorbing time-invariant user characteristics and common daily shocks. The vector $\mathbf{X}_i$ collects two pre-determined user characteristics interacted with Ban_t, which we describe in detail below. The interaction allows for parallel trends to hold conditional on $\mathbf{X}_i$ (Baker et al. 2026). The coefficient of interest is β_1, the triple interaction. We estimate the model via Poisson Pseudo-Maximum Likelihood (PPML) because Y_{it} is a non-negative count with substantial mass at zero; PPML accommodates these features while remaining consistent under arbitrary patterns of conditional heteroskedasticity.

FIGURE II. IMPACT ON USERS CONSUMING CONTENT FROM THE BANNED OUTLETS: DAILY EVENT-STUDY, TRIPLE-DIFFERENCE SPECIFICATION



Notes: The figure presents an event study for the triple-difference specification, estimated via Poisson Pseudo-Maximum Likelihood (PPML). The outcome variable is the daily count of pro-Russia tweets. Each point is the coefficient on the triple interaction between connectivity status, EU membership, and a day-specific indicator. We consider users connected if, before the ban, they interacted with the outlets directly or interacted with users who did. The omitted day is March 1^{st}, 2022, the day immediately preceding the implementation of the ban. The window runs from 8 days before to 13 days after the ban. The specification includes user and date fixed effects and controls for pre-ban attitudes and trade exposure with Russia, both interacted with the post-ban indicator rather than day-by-day, which produces virtually identical results while reducing the parameter count. Grey caps and hollow circles denote pre-treatment periods; violet caps and hollow squares denote post-treatment periods. Error bars are 95% confidence intervals. Standard errors are clustered at the user level. Flat and statistically indistinct pre-treatment coefficients support the parallel trends assumption.

The vector $\mathbf{X}_i$ contains two pre-determined, time-invariant user characteristics: the user's pre-ban share of pro-Russia content, capturing latent pro-Russia attitudes, and the trade exposure of the user's location to Russia, capturing local economic stakes in the conflict. Because both variables are time-invariant at the user level, they are mechanically absorbed by α_i and cannot be entered in levels. We therefore interact them with the post-ban indicator Ban_t, which allows users with different pre-ban attitudes and different local trade exposure to follow differential trajectories around the ban date. Including $\mathbf{X}_i \cdot Ban_t$ removes two natural confounds: that users with stronger pre-ban pro-Russia attitudes were on a steeper post-invasion trajectory, and that users in localities with stronger economic ties to Russia may have faced different information and incentive environments after 24 February 2022.

Table I reports estimates from three specifications. Column 1 contains only the saturated triple-difference; Columns 2 and 3 progressively add the interacted controls. The sample is fixed across columns at 1,141,514 user-day observations (the Column 1 sample after PPML drops always-zero users) to ensure that coefficient movement across columns reflects the addition of controls, not changes in sample composition. In our preferred specification (Column 3), $\hat{\beta}_1 = -0.245$ (s.e. = 0.072): connected EU users reduce their expected pro-Russia output by 21.7% post-ban ($(1 - e^{-0.245}) \times 100$). The estimate is stable between Columns 2 and 3, indicating that trade exposure does not materially confound the result conditional on pre-attitudes. The movement from Column 1 ($\hat{\beta}_1 = -0.148$) to Column 2 ($\hat{\beta}_1 = -0.259$) reflects a steeper post-invasion trajectory among users with stronger pre-ban pro-Russia attitudes: failing to account for this differential trend pulls the triple-difference toward zero. We revisit the two-way interactions β_2 and β_3 in Subsection 5.2 when we widen the lens to the population-wide effect of the ban.

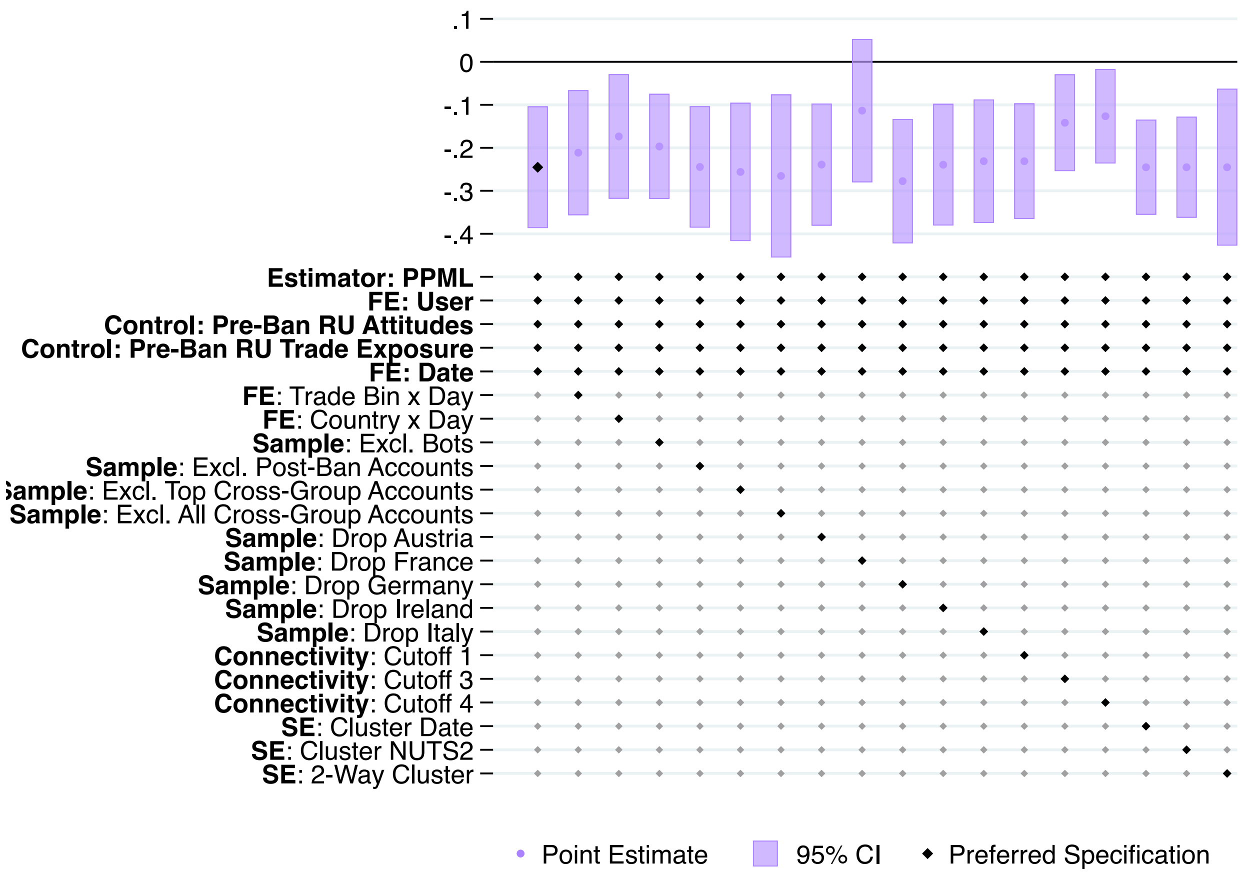
Figure II reports a daily event-study version of the triple-difference, plotting $\hat{\beta}_1$ separately for each day around the ban with March 1 as the omitted reference, the day before the ban took effect on the platform. The pre-ban coefficients are flat and statistically indistinguishable from zero from six days before the ban onward, supporting the parallel-trends assumption. The post-ban coefficients turn negative at the onset of the ban and stay negative across most days in the following two weeks, with a few near-zero days.

Taken together, these results point to a considerable impact of the ban among users connected to the banned outlets. As a back-of-the-envelope illustration, connected EU users posted on average 1.26 war-related tweets per day before the ban, of which roughly 4 in 10 (about 0.50 tweets/day) were pro-Russia. The 21.7% reduction, therefore, corresponds to about 0.11 fewer pro-Russia tweets per day per connected EU user. The event-study graph reinforces this picture: despite the inherent noise of daily social media data, the timing of the estimated effect aligns sharply with the ban date, with pre-ban coefficients flat and close to zero and a sudden drop at the cutoff that persists throughout the post-period. This sharpness is notable in itself. Related work on the social-media effects of content moderation policies (e.g., Müller and Schwarz (2022)) typically aggregates outcomes to the weekly or monthly level to recover sufficiently precise estimates. Detecting a cleanly identified effect at the daily level reflects both the abruptness of the ban's implementation and the strength of the response among directly connected users.

Robustness of Triple-Difference Specification Figure III reports the results of a series of robustness checks designed to assess the sensitivity of our estimates to alternative modeling choices. We re-estimate the preferred specification of Table I (Column 3) across 17 alternative PPML specifications that vary along four dimensions: the fixed effects structure, the estimation sample, the network-distance cutoff used to define a *connected* user, and the standard-error clustering choice. Each modeling choice is recorded in the grid below the coefficient plot, where an active marker indicates that the corresponding choice is in use for that specification. The grid should be read column by column: the full set of active markers for a given column describes the complete set of modeling choices underlying that estimate. The first coefficient reports the result of our preferred specification, from the main table of results.

The first set of robustness checks addresses the concern that countries in our sample may have experienced differential day-to-day shocks around the invasion, independent of the ban. While the triple-difference already absorbs any shock common to all EU users, it does not rule out heterogeneous daily dynamics across EU countries; France and Germany, for example, may have reacted differently on specific days for reasons unrelated to the ban. Two alternative fixed effects structures tackle this. The second

FIGURE III.
SUMMARY OF ROBUSTNESS CHECKS ON THE TRIPLE-DID SPECIFICATION



Notes: The figure reports point estimates and 95% confidence intervals for the triple interaction coefficient across 17 alternative specifications, all using the number of pro-Russia tweets as the dependent variable, estimated via Poisson Pseudo-Maximum Likelihood (PPML). The preferred specification – identified in the results above and including user and date fixed effects, pre-ban attitudes, and trade exposure as controls – is highlighted in black; all other estimates are shown in violet. The lower panel displays indicator dots recording which modeling choices are active in each specification, organized into four groups: fixed effects (date FE, trade-bin × day FE, country × day FE); sample restrictions (exclude bots, exclude post-ban accounts, exclude cross-interaction groups, leave-one-out for each of the five EU countries); connectivity cutoffs (cutoff 1, 3, and 4, with the baseline being cutoff 2; the baseline implies that we consider users as connected to the outlets if they directly interacted with the banned outlets, degree 1, or indirectly via one intermediary user in our sample, degree 2); and standard error clustering choices (by date, NUTS2 region, or two-way clustering by user and country-day).

model, from left to right in the figure, adds an interaction between each country's pre-ban trade exposure to Russia and a full set of day indicators, allowing countries with stronger economic ties to follow distinct daily trajectories. The third model goes further, interacting country identifiers directly with day fixed effects, absorbing any country-specific daily variation regardless of its source. In both specifications, the point estimate remains negative and statistically significant, indicating that the effect is not driven by day-to-day country-specific reactions.

The next set of checks modifies the estimation sample. We first address the concern that bot accounts could be driving the result. If Russian-linked bots were operating in the EU space before the ban and went quiet after it (because the ban disrupted the coordination network or eliminated two narrative sources), a post-ban drop in pro-Russia content could partly reflect a drop in automated activity rather than a behavioral change among human users. There is no universally accepted method for detecting bots in social media data, so we draw on an approach from the existing literature (Gehring and Grigoletto 2025): we classify as a bot any user who is in the top quartile of the tweet-production distribution and simultaneously in the bottom quartile of the reputation score (defined as the ratio of followers to the sum of followers and followees), or who produced more than 80 pro-Russia tweets in the eight days before the ban. This yields 2,585 users classified as bots, representing approximately 1.8% of the sample. Excluding them leaves the estimate essentially unchanged. The reverse concern, that EU-side bots ramped up pro-Russia activity to counteract the ban, would push our estimates toward zero, making the reported effect a lower bound on the ban's true impact on human users.

We next exclude accounts created after the ban date. The concern here is distinct from the bot check: new accounts could not have affected the pre-ban baseline, but if they were created post-ban specifically to amplify pro-Russia narratives in reaction to the ban, they would partially offset the measured decline among connected EU users and bias the triple-difference estimate toward zero. Excluding them, therefore, provides a direct test of whether such account creation is distorting our results. The estimate is essentially unchanged, indicating that newly created accounts play no meaningful role in our sample period. This is consistent with the short time window we consider: a coordinated, large-scale account creation campaign would likely take longer to organize and scale than the two weeks following the ban.

Next, we exclude accounts that may violate the stable unit treatment value assumption (SUTVA) in our setting. SUTVA requires that the ban's effect on one user does not depend on the treatment status of other users. A potential violation arises if EU and non-EU users communicate about the war with one another: information or content that circulates among non-EU users could reach EU users through direct mentions or replies, diluting the ban's effect and biasing our estimates toward zero. We proxy for this channel using a tweet-level indicator that flags tweets involving communication across the ban boundary, i.e., an EU account in our sample mentioning a non-EU account, or vice versa. We aggregate this indicator to the user level using pre-ban tweets only (to avoid endogeneity with the treatment), computing each user's share of pre-ban war-related tweets that are cross-group. The share of users with cross-group tweets in our sample is relatively low: only 7% of users exhibit cross-group pre-ban communication. We report two sample restrictions here: first, we exclude users whose share of pre-ban cross-group tweets exceeds 10%; second, we exclude users with any pre-ban cross-group activity. The results are comparable to our baseline in both magnitude and statistical significance; Appendix I.3 traces the full gradient of exclusions. This suggests that cross-ban spillovers do not strongly contaminate our estimates and that SUTVA violations are unlikely to be a primary driver of the main result.

With only five treated countries, a natural concern is that the triple-difference estimate is driven by the idiosyncratic dynamics of a single country rather than reflecting a broadly European response to the ban. We address this by re-estimating the preferred specification five times, dropping Austria, France, Germany, Ireland, and Italy in turn. In all five cases, the point estimate of β_1 remains negative, with values in the interval $[-0.28, -0.11]$. Confidence intervals exclude zero in all but the leave-France-out specification. France is the largest single source of connected users (and users in general) in our sample;

dropping it reduces the number of connected EU users substantially, widening the standard error enough for statistical significance to fade. The point estimate also attenuates in this case, to -0.11, so France contributes to the magnitude as well as the precision of the pooled estimate – unsurprising given its large share of connected users. The negative sign is nonetheless preserved across all five exclusions, and the other four leave the estimate close to baseline and significant, so the effect is not an artifact of any single smaller country.

The network-distance cutoff used to define a connected user is a discretionary choice in our design. We adopt degree-two as the baseline, motivated by the fact that users who interact with users who directly interact with the outlets are still genuinely embedded in the outlet-centered information network and meaningfully exposed to its content. We test three alternative cutoffs (degree-1, degree-3, and degree-4) to assess whether the results are sensitive to this choice. Restricting to degree-1 (users who directly interacted with the outlets) yields a coefficient virtually identical to the preferred degree-2 specification, indicating that second-degree users are also responding to the ban – not noise added to the connected group. Degree-3 and degree-4 produce estimates that are attenuated toward zero but remain statistically significant. This gradient is informative beyond robustness: if the effect were purely mechanical, it would be sharp and concentrated at degree-1. The smooth attenuation with network distance is instead consistent with an exposure and embeddedness mechanism, where users further from the network core were less reliant on the outlets' content and therefore less affected by their removal.

The final set of checks addresses inference. Our baseline clusters standard errors at the user level, but since EU membership is assigned at the country level, correlated errors within countries are a concern. With only seven countries, asymptotic justifications for country-level clustering are also weak. We test three alternatives – clustering by date, by NUTS2 region, and two-way by user and country-day – and find that the results remain statistically significant under all three. As a complementary check that requires no asymptotic justification, we conduct a within-country permutation test of the connected flag: randomly reassigning connection status within each country 100 times, none of the 100 permutations produces an absolute t-statistic as large as the one estimated in the actual data ($|t| = 3.39$; Appendix D, Table D.1).

5.2 Reduced-Form Effect of the Ban

The triple-difference in Section 5.1 documents the response to the ban among users most directly exposed to the RT/Sputnik network, users who interacted with the outlets directly or with others who did, relying on them to access Kremlin-aligned content to amplify and inspire their own posts. This effect is a necessary first stage for the supply-side interpretation of the wider response: it must hold for any aggregate decline among EU users to be attributed to disrupted outlet supply rather than to other channels such as signaling or norm shifts.

We now widen the lens to examine whether the ban's effects extend beyond the network of the banned outlets. Several channels from our conceptual framework (Section 3) could propagate the effect to unconnected users: a thinner overall supply of pro-Russia narratives in the platform's information environment, a lack of guidance on what content to push, or even self-censorship among users who fear similar moderation. We now move to a two-way difference-in-differences specification that captures the ban's reduced-form effect across the full EU user population, exploiting the geographical variation between EU and non-EU users and the timing of the ban's implementation.

To capture the ban's total reduced-form effect on the full EU user population, we now estimate a two-way difference-in-differences specification at the user-day level via PPML:

$$\log \mathbb{E}[Y_{it} \mid \alpha_i, \delta_t, EU_i \times Ban_t, \mathbf{X}_i] = \alpha_i + \delta_t + \beta\,(EU_i \times Ban_t) + \mathbf{X}_i'\boldsymbol{\gamma} \cdot Ban_t. \tag{2}$$

Y_{it}, EU_i, Ban_t, α_i, and δ_t are defined as in Section 5.1, and $\mathbf{X}_i'\boldsymbol{\gamma} \cdot Ban_t$ is the same vector of two pre-determined user controls (pre-ban pro-Russia attitudes and local trade exposure) interacted with the post-ban indicator. The coefficient β captures the average post-ban change in the outcome among EU users relative to non-EU users. As discussed in Section 4.3, this two-way DiD provides the aggregate reduced-form estimate across the full EU user population. The subsequent decomposition and mechanism analysis (from Section 6.1 to Section 6.6) then attribute the aggregate response to specific channels.[13]

Given the aim is to understand the generalized reduced-form impact of the ban, using the raw count of pro-Russia tweets as the sole outcome would provide an incomplete picture for three reasons. First, the estimates may be driven by a small set of highly active users whose behavior changed sharply after the ban, disproportionately influencing aggregate results. Second, absolute counts can mask meaningful compositional changes: a user who reduces output from fifteen pro-Russia tweets to fourteen has barely shifted, whereas one who goes from one to zero has stopped entirely. Third, the ban may affect not what users post about the war but whether they post about it at all, in which case a decline in pro-Russia content would reflect withdrawal from the topic rather than a genuine shift in stance. We address these three concerns through the four-column structure of Table II.

The table reports four outcomes that together characterize this broader response. Column 1 examines the extensive margin of pro-Russia posting, whether a user posts any pro-Russia content on a given day, and finds that treated users are 5.3% less likely to do so post-ban. Column 2 turns to overall posting activity and shows that total tweet volume falls modestly, by 6.7% (significant at the 10% level). The comparison suggests that reduced activity may contribute to the headline decline, but does not account for the overall pro-Russia specific decline: treated users are not primarily disappearing from the sampled war-related conversation. Column 3 delivers the main reduced-form estimate: total pro-Russia output among treated users falls by 13.6% relative to non-EU users. Column 4 restricts attention to user-days with at least one tweet and shows a decline of 11.0%, confirming that the Column 3 effect is not only a consequence of reduced activity but also reflects a compositional shift in what active users post.

Taken together, the four estimates describe a ban whose reduced-form effect operates primarily through the content that users post rather than whether they appear in the sampled conversation at all. Treated users remain active in the corpus at rates close to their pre-ban trajectory, but those who do post produce meaningfully less pro-Russia content. This is consistent with a ban that shifts the composition of observed war-related discourse on the platform, not merely one that removes a fixed category of content

[13] In Appendix E, we report event-study versions of Equation 2 that replace Ban_t with a full set of day-specific indicators Ban_t^k (k indexes days relative to the ban, $k = -1$ omitted as reference) to visualize pre-trends and the dynamic post-ban response.

TABLE II.
OVERALL IMPACT OF THE BAN IN EU VS. NON-EU COUNTRIES: DIFF-IN-DIFF

Dependent variable	**Any Pro-RU**	**Total Tweets**	**# Pro-RU Tweets**	
Margin	**Extensive**	**Volume**	**Unrestricted**	**Cond. on Active Day**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	-0.054 (0.019) [0.005]	-0.070 (0.036) [0.054]	-0.146 (0.032) [0.000]	-0.117 (0.027) [0.000]
Unconditional	✓	✓	✓	
Conditional on Posting				✓
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.137	0.577	0.207	0.757
Approx. Percentage Change	-5.28	-6.71	-13.63	-11.03
Observations	1141514	1141514	1141514	215723

Notes: The table presents two-way difference-in-differences estimates of the effect of the ban on users' posting behavior, estimated via Poisson Pseudo-Maximum Likelihood (PPML). The sample is the balanced user-day panel restricted to the post-invasion period: from 22^{nd} February to 15^{th} March. The coefficient of interest is the Ban × EU interaction. All columns include user and date fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. The outcome varies by column: Column 1 is an indicator for posting any pro-Russia content on a given day (extensive margin); Column 2 is the total volume of tweets posted about the war; Column 3 is the count of pro-Russia tweets on the balanced panel; Column 4 restricts the estimation to days with at least one tweet and captures compositional shift on active days. Columns 1, 2, and 3 share a common estimation sample: because all three specifications share an identical fixed-effect structure, PPML applies the same separation criterion and retains the same observations across all three. Column 4 runs on the active-day sample. Pre-ban means, and approximate percentage changes are computed on each column's own estimation sample. Percentage changes are $(e^{\beta} - 1) \cdot 100$. Standard errors are clustered at the user level..

from view. Appendix E reports the same four-outcome specification estimated separately on the connected and non-connected subsamples (Tables E.1–E.2 and Figures E.3–E.4). The decline is substantially larger among connected users than among non-connected users, consistent with the supply-disruption reading of the triple-difference. The 13.6% estimate in Column 3 bundles two conceptually distinct components: content directly referencing or linking to the outlets, and other pro-Russia content posted after the outlets became harder to access. The next subsection separates these within the same DiD framework, asking which component drives the headline estimate and what this reveals about the channel through which the ban operated.

6 Mechanisms and Channels

The triple-difference design explored above establishes a well-identified first-stage result: the removal of RT and Sputnik had a direct and measurable impact on users closely connected to the banned outlets, who sharply reduced their pro-Russia output in the days following the ban. This effect, however, extends to a wider audience: the two-way DiD reveals a significant decline across the full EU sample, suggesting the ban's reach went beyond the network immediately surrounding the outlets, and had a cascade effect among the general EU user population. The subsections that follow take up this broader result and, using the same DiD specification, examine the channels and mechanisms through which the ban operated.

6.1 Breaking Down the Impact: Outlet-Linked vs. Residual Content

The reduced-form estimates in Section 5.2 bundle together two conceptually distinct components. The first is *outlet-linked* content: pro-Russia tweets that retweet, share, mention, or link to RT/Sputnik content. Direct links, retweets, and visibility of the underlying accounts were restricted after the ban, although plain @-mentions of the outlet handles remained technically possible.[14] The second is *residual* content: pro-Russia tweets that do not cite the banned outlets. A decline in this residual category is consistent with behavioral adjustment, but it does not isolate a pure behavioral channel because users may paraphrase, post screenshots, copy frames, or relay content through intermediaries. Disentangling these components matters for policy interpretation: a ban that only reduces direct outlet citations has a narrower footprint than one associated with a broader change in other pro-Russia postings.

We explore these results in Table III. The decomposition is additive at the user-day level: $Y_{it}^{\text{total}} = Y_{it}^{\text{cite}} + Y_{it}^{\text{other}}$, where the first term collects pro-Russia tweets that mention or link to RT or Sputnik (*not* only direct retweets) and the second collects the residual. We estimate the two-way DiD specification on each component; Column 1 isolates the direct outlet-citation component, Column 2 isolates the residual posting response, and Column 3 reproduces the total effect from Section 5.2 for reference.

Column 1 shows that direct citations of RT and Sputnik fall by approximately 68% among treated users post-ban – the expected sign and a magnitude close to the upper bound of what a direct outlet-linked restriction of this type can deliver. The estimate is identified from only 356 users who posted at least one direct RT/Sputnik citation during the analysis window; users who never cited the outlets are absorbed by the user fixed effects and contribute nothing. Within this subset, the pre-ban baseline is 0.14 outlet-linked tweets per user-day. The 68% decline, therefore, reads as follows: among users with any pre-ban direct citations, the ban nearly eliminated this behavior.

Column 2 examines the residual posting response. Pro-Russia tweets that do not cite RT or Sputnik fall by 13.1% among treated users post-ban, estimated on the full panel. The magnitude is close to the total-effect estimate of 13.6% in Column 3, and the similarity is not a coincidence. At the population level, RT/Sputnik citations are a negligible share of aggregate pro-Russia output: the 356 users contributing to Column 1 represent less than one percent of the treated user base, and even among them, the pre-ban citation rate is modest. The residual component, by contrast, operates across the entire panel of users and on the overwhelming majority of pro-Russia content that is not a direct citation. The aggregate decline measured in Column 3 is therefore almost entirely attributable to Column 2 rather than Column 1.

This pattern has a concrete interpretation. The ban is not primarily associated with preventing

[14] Plain @-mentions of the outlet handles remain technically possible post-ban, since the EU measures geo-blocked broadcasting but did not remove the accounts (Section 2.3). We treat tweets that mention the outlets as a mixed case in the decomposition; the substantive results are unchanged when mentions are reassigned to the residual channel.

TABLE III.
DECOMPOSITION: OUTLET-LINKED VS. RESIDUAL PRO-RUSSIA CONTENT

Dependent variable	**No. Pro-Russia Tweets Posted**		
Component	**Outlet-linked**	**Residual**	**Combined**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Ban × EU	-1.145 (0.364) [0.002]	-0.140 (0.032) [0.000]	-0.146 (0.032) [0.000]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.138	0.207	0.207
Approx. Percentage Change	-68.18	-13.09	-13.63
Observations	7832	1139776	1141514

Notes: The table decomposes the overall effect of the ban on pro-Russia content production into outlet-linked and residual components, both estimated via Poisson Pseudo-Maximum Likelihood (PPML). The coefficient of interest is the Ban × EU interaction. All columns include user and date fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. Column 1 estimates the effect on *RT/Sputnik outlet-linked* pro-Russia content (tweets that mention or link to RT or Sputnik, not only direct retweets). Column 2 estimates the effect on residual pro-Russia content (not RT/Sputnik-linked). This residual response is consistent with behavioral adjustment, but does not isolate a pure behavioral channel: users may paraphrase, post screenshots, copy frames, or relay content through intermediaries. By construction, the two components sum at the user-day level to the total measure in Column 3, which repeats the main specification on all pro-Russia tweets. The decomposition is at the tweet level, thus a single user may contribute to both columns. Because PPML estimates multiplicative effects, the percentage changes in Columns 1 and 2 do not sum arithmetically to Column 3. Column 1 is identified from the subset of users who had any RT/Sputnik-linked pro-Russia content in the analysis window; users who never did are absorbed by the user fixed effects. Because Column 1 is estimated on this small, selected subset, its magnitude is not directly comparable to the population-wide estimates in Columns 2 and 3. Pre-ban means and percentage changes are computed on each column's estimation sample. Standard errors are clustered at the user level.

a fixed set of directly linked outlet content from reaching European audiences; while the outlet-linked component exists and is sharp where it applies, it operates on a negligible fraction of pro-Russia output. Most of the reduced-form decline appears in residual pro-Russia content. Users who continue to post on Twitter after 2 March produce meaningfully less pro-Russia content, even when that content does not directly reference the banned outlets and could still technically be posted.

The residual component accounts for nearly all of the aggregate decline, a finding that is not surprising given that outlet-linked tweets represent a negligible share of total pro-Russia output. But the implication is broader: the removal of a relatively small set of outlet-linked content was associated with a much broader contraction in residual pro-Russia posting across the full user population. This points to the banned outlets having played a role that extended well beyond their direct citation footprint. Two questions follow naturally. First, did other users attempt to fill the void left by the outlets, stepping up their own pro-Russia output to compensate for the decline? Second, through what mechanism did the ban reduce posting even among users with no direct connection to the outlets? We tackle these questions in the sections below.

6.2 Impact on Pre-Ban Pro-Russia Suppliers

The evidence we presented so far establishes that the ban created a real supply shock: EU-based users produced fewer pro-Russia tweets in absolute terms, were less likely to post such content on a given day, and this decline extended well beyond the network of users directly connected to the banned outlets. A void of pro-Russia content formed in the Twitter discourse of our sample. The natural question is whether other users stepped in to fill it: did committed producers of pro-Russia content, directly unaffected by the ban, exploit the gap to increase their own output and establish themselves as more prominent suppliers?

We define *Pro-Russia producers* as those users who created at least one pro-Russia tweet – original or reply, excluding retweets – in the eight days before the ban. This population spans both connected and unconnected users, according to whether they interacted with the banned outlets pre-ban. Within it, we focus on *secondary suppliers*: producers active on at least four of those eight days, for whom pro-Russia posting is a sustained behavior. The "supplier" label captures the role this group plays in the sampled discourse: at a smaller scale and without an editorial apparatus, they generate a regular flow of pro-Russia content alongside the primary outlets (RT and Sputnik).

Secondary suppliers are the natural focus: unlike incidental producers (those active on fewer than four days), they were active participants in the dissemination of the content targeted by the ban, and thus the most likely to try to compensate for the outlets' removal. Appendix Figures B.4 and B.5 trace the full activity gradient (1 to 6+ days) across all pro-Russia producers and motivate the four-day threshold: network proximity to the outlets, pre-ban engagement on pro-Russia content, and outlet-linked posting all concentrate at higher activity levels.

As a first step, we reproduce the four-outcome specification from Section 5.2 on the supplier subsample. Table IV reports the results. The pattern is unambiguous in two ways. First, suppliers do not compensate for the ban within our sampled corpus: rather than ramping up their pro-Russia output to fill the gap left by RT and Sputnik, they reduce it. Column 3 shows that total pro-Russia output among treated suppliers falls by 14.5%, a magnitude close to the aggregate effect in the full sample. Second, the decline is concentrated on the intensive margin. Column 1 shows that the probability of posting any pro-Russia content on a given day is essentially unchanged ($\hat{\gamma} = -0.007$, $p = 0.88$, against a pre-ban probability of 0.63): suppliers are not less likely to post any pro-Russia content on a given day. Column 4 conditions on days with at least one tweet of any kind and recovers a 16.4% reduction in pro-Russia output – the largest compositional effect we observe on any margin. Column 2 shows that overall tweet volume falls by 9.4% (marginally significant, $p \approx 0.10$), larger than the 6.7% decline observed for the average EU user, indicating that suppliers also tweet less in general post-ban, but not enough to mechanically explain the pro-Russia decline.

The simplest substitution story does not survive the test of data exploration: the users best positioned to replace the banned outlets reduce rather than expand their pro-Russia output. A natural next question is whether this aggregate decline is broad-based or concentrated in a particular subgroup of producers.

TABLE IV.
IMPACT OF THE BAN ON PRE-BAN PRO-RUSSIA SUPPLIERS: DIFF-IN-DIFF

Dependent variable	**Any Pro-RU**	**Total Tweets**	**# Pro-RU Tweets**	
Margin	**Extensive**	**Volume**	**Unrestricted**	**Cond. on Active Day**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	-0.007 (0.047) [0.879]	-0.099 (0.060) [0.100]	-0.157 (0.076) [0.039]	-0.179 (0.076) [0.019]
Unconditional	✓	✓	✓	
Conditional on Posting				✓
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.628	3.570	1.751	2.323
Approx. Percentage Change	-0.71	-9.44	-14.54	-16.37
Observations	29854	29854	29854	17053

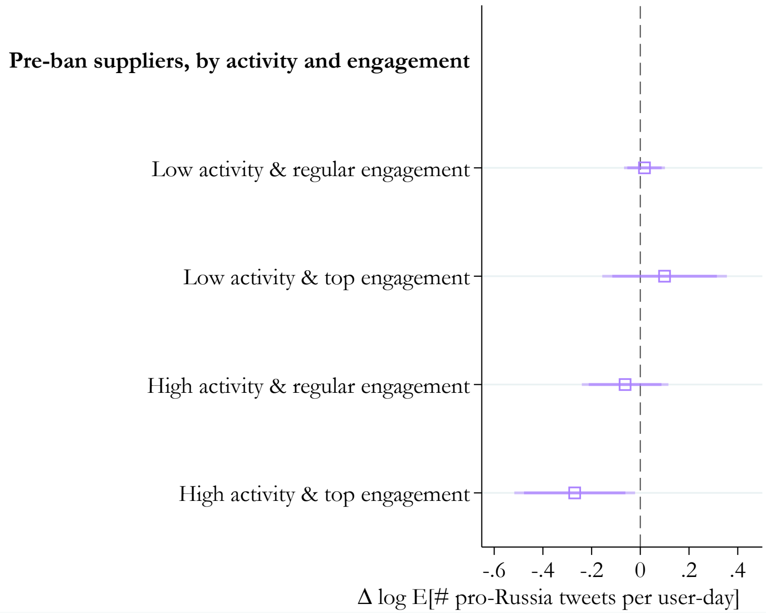
Notes: The table re-estimates the specification of Table II on the subsample of *pre-ban pro-Russia secondary suppliers*: users who posted pro-Russia content on at least four of the eight pre-ban days in the analysis window. All the columns show two-way difference-in-differences estimates of the effect of the ban on users' posting behavior, estimated via Poisson Pseudo-Maximum Likelihood (PPML). The sample is the balanced user-day panel restricted to the post-invasion period (from 22$^{\text{nd}}$ February to 15$^{\text{th}}$ March) and to secondary suppliers only. The coefficient of interest is the Ban × EU interaction. All columns include user and date fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. The outcome varies by column: Column 1 is an indicator for posting any pro-Russia content on a given day (extensive margin); Column 2 is the total volume of tweets posted about the war; Column 3 is the count of pro-Russia tweets on the balanced panel; Column 4 restricts the estimation to days with at least one tweet and captures compositional shift on active days. Columns 1, 2, and 3 share a common estimation sample: because all three specifications share an identical fixed-effect structure, PPML applies the same separation criterion and retains the same observations across all three. Column 4 runs on the active-day sample. Pre-ban means and approximate percentage changes are computed on each column's own estimation sample. Percentage changes are $(e^{\beta} - 1) \cdot 100$. Standard errors are clustered at the user level.

We address this in the next subsection.

6.3 Heterogeneous Effects on Producers of Pro-Russia Content

We now broaden the sample to the full population of pre-ban pro-Russia producers, defined as users who posted at least one original pro-Russia tweet or reply before the ban. The analysis above focused on the most committed subset of this population: users active on at least four pre-ban days, whom we defined as secondary suppliers. Here we ask which subgroups of producers were most affected by the ban. We examine two dimensions of heterogeneity: pre-ban activity level and pre-ban engagement. For activity, we distinguish low-activity producers (one to three days of pro-Russia posting over the eight-day pre-ban window) from high-activity producers (four or more days, the secondary supplier threshold). For engagement, we define high-engagement producers as those whose pro-Russia content ranked in the top 5% of the pre-ban distribution of retweets, replies, and likes per tweet.

FIGURE IV.
IDENTIFYING THE USERS MOST AFFECTED BY THE BAN



Notes: The figure displays coefficients and 90% (dark violet) and 95% (light violet) confidence intervals from four separate estimations of Equation 2, each restricted to a distinct subgroup of pre-ban pro-Russia producers, estimated via Poisson Pseudo-Maximum Likelihood (PPML). The dependent variable is the number of pro-Russia content tweets. The sample covers all user-day observations where the user posted at least one tweet about the war, over the period February 22nd–March 15th. The four subgroups are defined by crossing two dimensions: pre-ban activity (low: one to three active days; high: four or more, corresponding to the secondary supplier threshold) and pre-ban engagement (top: users whose pro-Russia content ranked in the top 5% of the pre-ban distribution of retweets, likes, and replies per tweet; regular: all others). The dependent variable is the daily count of pro-Russia tweets per user; raw coefficients capture changes in the logarithm of the conditional expectation. Percentage changes, computed as β as $e^{\beta} - 1$, and observation counts are reported in Appendix Table H.3. All specifications include user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level.

Figure IV presents the ban's effect on pro-Russia content production across the four subgroups. The pattern is sharp. Among low-activity producers, those below the secondary supplier threshold, the ban had no detectable effect, regardless of their pre-ban engagement level. This is consistent with these users being incidental producers of pro-Russia content, unlikely to have relied on RT and Sputnik as agenda-setters. Among high-activity producers, the picture is more nuanced: the overall decline documented above is driven entirely by those whose content attracted high pre-ban engagement. Highly active producers with regular engagement show no significant response; it is specifically the high-activity, high-engagement subgroup that accounts for the aggregate decline.

The concentration of effects among high-activity, high-engagement producers suggests two complementary interpretations. First, these users were most likely to depend on RT and Sputnik as a source of ready-made narratives: sustaining a high volume of pro-Russia content is demanding, and the outlets reduced that cost by supplying framings and topics that could be repurposed at low effort. Second, these users were also those whose content attracted the highest pre-ban engagement, making them visible in a context where the EU had just banned the most prominent pro-Russia voices. The possibility that enforcement could extend to the next tier of visible suppliers may have induced strategic self-censorship. While both interpretations are speculative, they are consistent with the observed pattern and lead to a follow-up question: did the broader population of users – those who consumed rather than produced pro-Russia content – redirect their attention to secondary suppliers once the primary outlets were removed?

6.4 Suggestive Evidence on the Demand Side

The analysis so far has examined the supply side: with two major suppliers removed, other users produce less pro-Russia content. We now turn to the demand side. Given the short time window, demand for pro-Russia content may be roughly inelastic in the days after the ban. If so, did consumers of pro-Russia content redirect their attention to alternative sources?

Figure V examines the effect of the ban on the per-tweet engagement that producers of pro-Russia content receive, where engagement is the number of retweets, replies, and likes received by each pro-Russia tweet. To connect this demand-side analysis with the supply-side heterogeneity results above, we use the same four-group classification as in Figure IV: producers divided by pre-ban activity level (low: one to three days; high: four or more days) and pre-ban engagement (regular; top 5%). Using the same classification allows direct comparison across the supply and demand sides of the market.

The figure reports results from four separate PPML regressions where the dependent variable is per-tweet engagement – the average number of likes, retweets, and replies received per pro-Russia tweet posted by the user. The estimates are imprecise: in three of the four subgroups they are statistically indistinguishable from zero, and the lone exception – a positive estimate for low-activity producers with regular engagement – is significant only at the 10% level, not at the 5%. Crucially, none of the high-activity producer groups, those best positioned to absorb the audience left by the banned outlets, show any significant change in per-tweet engagement after the ban.

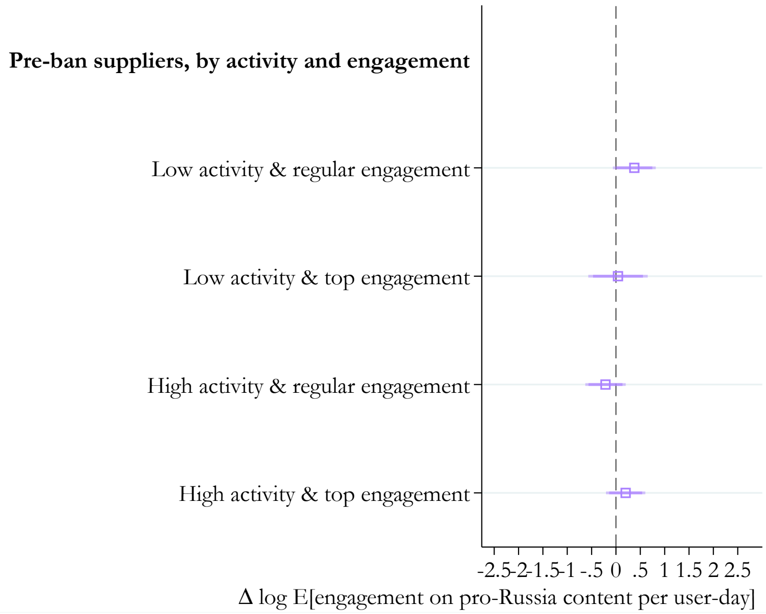
Thus, we find no evidence that engagement within the sampled corpus systematically redirects from the banned outlets to the alternative producers most able to absorb it. Per-tweet engagement on pro-Russia content shows no statistically significant change at the 5% level for any subgroup after the ban – neither the most active producers nor those with the highest pre-ban audiences attract more attention in the absence of RT and Sputnik. Combined with the supply-side finding that secondary suppliers reduce rather than increase their pro-Russia output, the evidence points to a discourse in which neither supply nor observed audience engagement compensates for the loss of the banned outlets.

While we cannot rule out substitution outside the sampled corpus or on other platforms, the evidence within it points to a short-term contraction of pro-Russia content with no compensating response on either the supply or the demand side. We examine one further substitution channel – institutional reallocation by other Russian state-backed sources – before turning to the mechanism behind the contraction.

6.5 Institutional Substitution: The Russian News Agency (TASS)

Beyond substitution within the Twitter user base, we examine whether the ban prompted a substitution by other Russian state-backed sources. The European institutions justified the ban by arguing that, despite their claimed independence, Russia Today and Sputnik acted as amplifiers of the Russian state narrative.

FIGURE V.
IDENTIFYING THE USERS MOST AFFECTED BY THE BAN: ENGAGEMENT



Notes: The figure displays coefficients and 90% (dark violet) and 95% (light violet) confidence intervals from four separate estimations of Equation 2, each restricted to a distinct subgroup of pre-ban pro-Russia producers. Throughout, pro-Russia content refers to original tweets and replies, excluding retweets. The sample covers all user-day observations where the user posted at least one tweet about the war, over the period February 22$^{\text{nd}}$–March 15$^{\text{th}}$. The four subgroups are defined by crossing pre-ban activity (low: one to three active days; high: four or more, corresponding to the secondary supplier threshold) and pre-ban engagement (top: users whose pro-Russia content ranked in the top 5% of the pre-ban distribution of retweets, likes, and replies per tweet; regular: all others). The dependent variable is per-tweet engagement – the average number of likes, retweets, and replies – received per pro-Russia tweet posted by the user. Raw coefficients capture changes in the logarithm of the conditional expectation; percentage changes (β as $e^{\beta} - 1$) and observation counts are reported in Appendix Table H.4. All specifications include user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level.

If the institutional resources behind that role were redirected elsewhere once the outlets were removed, we should see a footprint in the activity and engagement of other state-linked accounts. Appendix G examines this by comparing TASS – the Russian state news agency – to RT and Sputnik before and after the ban.

The point estimates do not support an institutional-substitution interpretation. Pro-Russia output by TASS, in relative terms, declines rather than rises after the ban, and per-tweet engagement with its pro-Russia content rises only modestly. None of the differences is statistically distinguishable from zero, so we cannot rule out small adjustments either way. Two readings are consistent with this null. First, TASS may primarily function as a wire service producing short news updates rather than the sustained, framing-rich social-media propaganda that RT and Sputnik specialized in – so it is poorly placed to substitute for the banned outlets even if institutional attention had shifted. Second, the two-week post-ban window is short for a coordinated institutional reallocation. Either way, our data show no evidence that the void left by RT and Sputnik was filled by another major Russian state-backed source on the platform.

6.6 Russia Today and Sputnik as Agenda-Setters

The results we reported so far are consistent with the interpretation of RT and Sputnik as agenda-setters, as we formulated in the conceptual framework above. Their ban led to a reduction of ready-made narratives that constituted a supply shock for users – directly and indirectly dependent on that content – and did not generate a short-term substitution by secondary suppliers. In this final step we provide the most direct evidence of this agenda-setting interpretation by examining the topical content of tweets in our sample: the outlets' daily output defined specific narrative frames that EU pro-Russia users tracked, and their removal disrupted that synchronization.

We examine two complementary dimensions in Table V: topical diversity (Columns 1 and 2) and topical alignment with the banned outlets (Columns 3 and 4). Topical diversity is the number of distinct topics covered per tweet, computed across all content (Column 1) or within pro-Russia content specifically (Column 2). Topical alignment is the share of a user's tweets on day t that cover topics appearing in RT and Sputnik's top five topics on the same day, across all tweets (Column 3) and within pro-Russia tweets specifically (Column 4). The outlet top-5 is dynamic, computed day by day from the outlets' own tweets, and captures the specific narrative frames the outlets pushed in real time rather than a fixed set of stable themes.

Columns 1 and 2 find essentially no effect. Users' overall topical diversity is unchanged post-ban ($\hat{\gamma} = 0.000$, $p = 0.995$, against a pre-ban baseline of 2.80 topics per tweet), and the same is true for their topical diversity within pro-Russia content ($\hat{\gamma} = 0.013$, $p = 0.55$, baseline 3.25). These are informative nulls: the estimates are small relative to the pre-ban means. The ban does not cause users to narrow or broaden the range of subjects they discuss.

Column 3 tells a different story. The share of treated users' tweets covering RT and Sputnik's daily top five topics falls by roughly 17% post-ban ($\hat{\gamma} = -0.014$, $p < 0.001$, against a pre-ban baseline of 0.081). Because Column 3 normalizes by all tweets, this overall decline mixes two channels: lower pro-Russia tweet volume and compositional shifts within remaining pro-Russia activity. Column 4 isolates the latter by measuring the share of pro-Russia tweets covering the outlets' daily top-five topics on days with at least one pro-Russia tweet; the conditional decline is 11% ($\hat{\gamma} = -0.013$, $p = 0.068$, against a pre-ban baseline of 0.117), suggesting that the alignment shift operates within the pool of pro-Russia tweets rather than purely through volume. Users are not covering fewer topics, but the topics they cover are less synchronized with the banned outlets' day-to-day priorities.

This pattern has a natural interpretation. Before the ban, EU users interested in pro-Russia content could observe RT and Sputnik's output directly and in real time. The outlets' daily editorial choices – which stories to emphasize, which frames to push, which themes to lead with – defined a live agenda that other users could track. After the ban, this real-time observation became much harder. Even motivated users who continued producing pro-Russia content were less able to mirror an agenda they could no longer see directly. The result is a systematic drift in topical content: users continue to cover diverse subjects, but lose synchronization with the narratives the banned outlets would otherwise have been promoting on a given day.

Three findings line up to support an agenda-setting reading of the ban's effect. Section 6.1 shows

TABLE V.
IMPACT OF THE BAN ON PRO-RUSSIA CONTENT TOPICS

Dependent variable	**Topics Rate in All Tweets**	**Topics Rate in Pro-RU Tweets**	**Top-5 Overlap (All Tweets)**	**Top-5 Overlap (Pro-RU Tweets)**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	0.000	0.013	-0.014	-0.013
	(0.012)	(0.022)	(0.003)	(0.007)
	[0.995]	[0.547]	[0.000]	[0.068]
Conditional on Posting	✓		✓	
Conditional on Posting Pro-Russia		✓		✓
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome for Treated	2.796	3.251	0.081	0.117
Change as % of Pre-Ban Mean	0.00	0.40	-17.09	-11.30
Observations	322199	75247	322199	75247

Notes: The table examines the effect of the ban on topic content, where topics are classified by our LLM-based pipeline, which extracts up to five topics per tweet. All four columns are estimated via two-way fixed effects OLS. The coefficient of interest is the Ban × EU interaction. All specifications include user and date fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. Outcomes across columns: Column 1 is the average number of topics per tweet (all content) on days with at least one tweet; Column 2 is the average number of topics per pro-Russia tweet on days with at least one pro-Russia tweet; Column 3 is the share of a user's tweets whose topics overlap with RT and Sputnik's five most-posted topics on the same calendar day, computed on days with at least one tweet (all-tweet denominator); Column 4 is the corresponding share among pro-Russia tweets, computed on days with at least one pro-Russia tweet. The outlet top-5 is computed dynamically day by day from the outlets' own tweets, not fixed across the window. Pre-ban means and approximate percentage changes are computed on each column's own estimation sample. Because the outcomes are rates or shares rather than count variables with mass at zero, we compute the percentage change as $\hat{\beta}/\bar{y}_{\text{pre}}$ rather than $e^{\beta} - 1$. Standard errors are clustered at the user level.

that the population-level decline is concentrated in residual pro-Russia posts rather than in direct outlet citations: most of the contraction is in content that does not name the outlets. Section 6.4 shows no compensating shift on the demand side, ruling out a story in which audiences simply moved to alternative producers. The topical-alignment evidence in this section then provides direct support for the mechanism: users continue to cover war-related subjects but lose synchronization with the narrative frames the banned outlets would have been pushing. Together, this is consistent with the outlets having functioned as agenda-setters for a connected network of users who tracked their daily framings, with the 17% drop in topical alignment serving as one empirical signature of that role. Appendix J provides complementary evidence on a specific class of narratives. Linking tweets in our sample to the 26 disinformation cases published between February 1 and March 31, 2022, concerning the war in Ukraine, by EUvsDisinfo, the EU's dedicated tracker of disinformation campaigns, using both semantic similarity and keyword matching, we find consistent negative effects across all ten measures post-ban. This suggests that the disruption to RT/Sputnik's agenda-setting role might have extended to the propagation of documented disinformation campaigns, not only to the broader topical alignment we measure here.

7 Threats to Analysis

This section reports robustness checks and placebo tests for the observational results; full evidence is in the appendices. The specification chart in Figure III already provides a comprehensive robustness assessment of the triple-difference design, varying controls, sample restrictions, clustering choices, and outcome definitions. The results discussed in this section complement that exercise by exploring further specifications and settings for both the triple-difference and the difference-in-differences design.

Ruling Out Generalized EU Shocks. A natural concern with any EU vs. non-EU comparison around February–March 2022 is that EU countries experienced contemporaneous shocks unrelated to the ban – the Temporary Protection Directive of 4 March, the refugee inflows and humanitarian exposure that followed, and broader shifts in European public sentiment toward Russia – that may have depressed pro-Russia posting on their own. The triple-difference design partly addresses this by absorbing any change shared by all EU users, and our preferred specifications include trade exposure interacted with the ban to proxy for country-level economic exposure to the war. The specification chart (Figure III) further shows that adding country-by-day fixed effects – which absorb any time-varying country-level shock – reduces the magnitude of the estimate somewhat but leaves a meaningful and statistically significant negative effect.

Beyond these design features, the data themselves point away from a generalized EU-wide attitudinal shift. Figure E.5 estimates the ban's effect separately for users closely connected to RT and Sputnik in the pre-ban network (within two degrees) and for users further removed (three degrees or more). The reduction in pro-Russia output is substantially larger for the closely connected group. A uniform EU-wide shift in sentiment would affect both groups proportionally, regardless of their proximity to the banned outlets; the close-versus-far asymmetry is consistent with a supply-side disruption operating through the outlets themselves. Additionally, as we control for pre-ban pro-Russia attitudes, the asymmetry cannot be created by stronger pro-Russia leanings of the connected users.

Robustness Checks: Our first check concerns the LLM-based pro-Russia classification. As described in Section 4.2, we validate the measure with an embedding-based continuous slant constructed from tweets by Russian and Ukrainian government accounts. Appendix C replicates the main event studies for the triple-difference and DiD specifications using this alternative measure; our main findings – a sharp post-ban decline in pro-Russia content among EU users, with the largest decline among connected users – are qualitatively unchanged.

We address the impact of potential bots on our results. The specification chart already shows that excluding potential bots from the triple-difference makes no substantive change. Appendix I.1 complements this by reproducing the main event study from the triple-difference specification and the main table from the DiD specification with bots removed. Excluding these users does not substantively alter our headline estimates. Moreover, if Russian bots had attempted to counteract the effects of the ban, removing them from the sample would tighten rather than dilute the estimated decline, so our main estimates are conservative on this dimension. Appendix I.2 replicates the analysis after excluding users whose accounts were created only after the ban took effect, with no substantive change in the findings.

We next examine the sensitivity of our results to the estimator. The paper uses Poisson Pseudo-Maximum Likelihood (PPML) for count outcomes with substantial mass at zero, following Chen and Roth (2024). Appendix I.4 re-estimates the four-outcome two-way DiD using OLS. The OLS estimates reproduce the PPML findings on sign and significance: the effects on the extensive margin (Column 1), total tweet volume (Column 2), and total pro-Russia output (Column 3) are all negative and statistically significant, though the linear model attenuates the magnitudes on the skewed count outcomes. Only the conditional-on-active-day margin (Column 4) diverges, collapsing to near zero and statistical insignificance, consistent with OLS being noisy for count outcomes with mass at zero and reinforcing our preference for PPML.

Placebo Tests: We check whether the EU-specific decline could reflect a broader global trend rather than the ban itself. Appendix Figure E.1 shows that pro-Russia content drops sharply and persistently for EU users right at the ban date, with no comparable break for non-EU users.

We then assign a fictitious ban date five days earlier than the actual implementation and re-run the event study. Confounding factors unrelated to the ban would generate a spurious differential effect at this earlier date. Appendix Figure F.1 shows no such effect.

The ban targeted pro-Russia outlets and should not affect anti-Russia content. Appendix Figure F.2 confirms this: no systematic change in anti-Russia content following the ban.

We then rule out a broader shift in European media coverage around the ban. Using GDELT article-level data, Appendix F shows that EU-versus-non-EU newspaper tone exhibits no differential change at March 2, 2022 (Figure F.3, Table F.1). A 2014 Crimea-annexation placebo (Table F.2) is also null. The pro-Russia decline we document is Twitter-specific rather than reflecting a wider shift in how European media covered the war.

Spillovers and SUTVA: A final identification concern is that EU and non-EU users in our sample are not isolated, and stable unit treatment value assumption (SUTVA) violations could bias our estimates in either direction. In one direction, when EU users posted less pro-Russia content post-ban, non-EU users who followed them passively received a thinner pro-Russia diet in their feeds. If non-EU users responded by posting less themselves, the control group partly absorbs the treatment effect, biasing both estimators toward zero. In the other direction, the ban may have caused RT/Sputnik to reallocate content toward unrestricted non-EU markets, increasing non-EU pro-Russia activity post-ban. If the control group trends upward for this reason, the DiD comparison overstates the causal effect on EU users, biasing estimates away from zero. The net direction is theoretically ambiguous.

To assess the quantitative severity of this concern, we identify users with pre-ban cross-network activity, defined as any tweet involving communication across the ban boundary (an EU user mentioning a non-EU user, or vice versa). The specification chart (Figure III) shows that the triple-difference is robust to dropping these users. Appendix I.3 traces the population-wide DiD across an increasingly aggressive gradient of cross-network exclusions—dropping users whose pre-ban cross-boundary share exceeds 25%, then 10%, and finally any positive share (Figure I.3 and Table I.3). The estimate is stable throughout and remains statistically significant even under the most aggressive cut: the effect on pro-Russia output moves only from -13.6% at baseline to -11.2% when every cross-network user is removed, and the active-day margin is similarly flat (both significant at the 1% level). Because excluding any user with even a single cross-boundary tweet mechanically removes the most active producers, this mild attenuation is expected and does not indicate a spillover artifact. We conclude that cross-boundary spillovers are unlikely to drive the population-wide effect.

8 Does the Ban Carry a Cost?

The observational analysis shows that the ban substantially reduced the circulation of pro-Russia narratives across the EU. The intervention nevertheless raises a question that observational data on supply alone cannot answer: did censoring RT and Sputnik as a counter-disinformation tool come at a cost in the European democratic setting? Press freedom and media independence are widely viewed as constitutive of liberal democracy, and the large-scale removals of media outlets is in opposition with those norms, even when motivated by security concerns.

The Council's regulation was an unprecedented measure. The ban covered every distribution medium – cable, satellite, streaming platforms, and social media – making it one of the strictest actions ever adopted against a media organization in the EU (Popović 2022). While the speed of reaction can be read as a signal that democracies can act decisively in a crisis, the decision was adopted through informal channels, bypassing standard democratic mechanisms and overriding the principle that freedom of speech and media regulation are matters decided at the country level (Vériter 2025). This tension between security imperatives and democratic norms is not new: democracies that rely on emergency restrictions on speech risk creating precedents that erode the very foundations they claim to protect (Linz 1978).

The EU ban was framed as censorship in real time by those most directly concerned. The European Federation of Journalists, the continent's largest journalists' association, issued a statement the day before the ban took effect titled *Fighting Disinformation with Censorship Is a Mistake*, warning that the measure represented "a complete break with democratic guarantees" and that "this act of censorship can have a totally counterproductive effect on the citizens who follow the banned media" (Journalists 2022). This raises a direct empirical question: if citizens register the ban as censorship rather than as a proportionate democratic response, does their trust in free-speech protections suffer?

In this section, we examine suggestive evidence on precisely this question. Specifically, we ask whether making the EU's decision to ban Russia Today and Sputnik salient affects perceived EU commitment to free speech and media independence. The experiment provides additional evidence on a dimension the observational analysis cannot directly reach: whether the use of censorship comes at a measurable cost to public trust in free-speech protections.

Before turning to the design, we flag two features of what this experiment can and cannot identify. First, it measures responses in 2025 to a treatment that makes the 2022 ban salient – not the immediate response to the ban itself in 2022. Second, the treatment shifts salience rather than introducing information from scratch. We do not measure prior awareness of the ban, but to the extent that some respondents already had it in mind when answering, the experimental contrast understates the response to making the issue salient. Our estimates should be read as intent-to-treat magnitudes of a 2025 salience treatment, biased toward zero by any pre-existing awareness or active recall. A further interpretive question – whether responses reflect rational updating about EU free-speech protections or shifts in underlying trust – has different normative implications and we return to it after presenting the results.

Experimental Design: In July 2025, we recruited 900 participants through the online panel survey company Prolific. The sample includes 300 respondents from each of the three most populous EU countries: Germany, France, and Italy.[15] At the start of the survey, we informed participants that the study concerns their personal views on the EU's response to the Russia–Ukraine war. First, before treatment, we ask all participants to express their general satisfaction with how democracy works in the EU; this pre-treatment item is the baseline satisfaction-with-democracy covariate used in the analysis. Participants then read brief information about the war.

All participants receive two short information briefs about EU policies supporting Ukraine in the ongoing Russian–Ukrainian conflict: one on a humanitarian aid package and another on financial assistance for Ukraine. We randomly assign half of the participants to receive an additional brief on the 2022 ban on all broadcasting activities by Russia Today and Sputnik, the natural experiment that we investigate in this paper.[16] The treatment either informs participants about the ban if they are previously unaware of it, or increases the salience of the issue if they already know. We do not observe prior awareness of the ban in either group, but we hypothesize that the treatment increases its salience when participants answer subsequent questions about perceptions of core democratic institutions. To ensure that participants read and understand the informational briefs, we include a simple comprehension question and make survey completion contingent on a correct answer.

Once participants read the information briefs, they proceed to the section of the survey that explores their views on core democratic norms in the EU. Our primary, preregistered outcome is an index of freedom of speech, calculated as the average of two main items: agreement with the statement "The European Union protects freedom of speech" and the media-independence item, which was elicited in reverse wording and coded for analysis so that higher values indicate stronger perceived EU protection of media independence. We use a 7-point Likert scale to measure agreement with the statements and rescale the analysis variables so that higher values indicate stronger perceived EU commitment to free speech, media independence, or democratic protections. We also include two additional statements that capture a broader understanding of democratic norms, whose average forms an index that we preregistered as an exploratory outcome. To reduce the likelihood of experimenter demand effects, we ask filler questions about humanitarian aid, financial support for Ukraine, refugee policy, and sanctions, which mask the purpose of the study. As secondary outcomes, after treatment, we also measure trust in the national

[15] Because the sample is not representative, we include a comprehensive set of sociodemographic controls in our analysis, as specified in our AsPredicted pre-registration. We collected these variables at the end of the online survey.

[16] See Appendix K.2 for the full information briefs and all screens used in the online experiment.

government, the national parliament, the European Union, and the European Commission, as well as a separate post-treatment satisfaction-with-democracy item. The survey concludes with questions on sociodemographic characteristics. Appendix K.2 provides the full questionnaire and all items.

Experimental Evidence: We focus on the main outcome of our experiment: the index of freedom of speech. Appendix Figure K.1 shows that sociodemographic characteristics and baseline satisfaction with democracy in the EU are balanced across the two experimental conditions. As preregistered, we nevertheless control for sociodemographic characteristics in all analyses, given that the sample is not representative. In addition to personal characteristics, we include language fixed effects – participants could choose to take the survey in English, German, French, or Italian – and we control for baseline satisfaction with democracy in the EU. Table VI reports the main findings of our experiment. The table presents coefficients from simple OLS models estimating the effect of the treatment information on the index of freedom of speech (Column 1) and on the two separate components of the index: agreement with the statement on freedom of speech and agreement with the statement on media independence, respectively, in Columns 2 and 3.

We find that perceptions of freedom of speech are lower in the treatment group that receives information on the ban. The effect is statistically significant at the 10% level (two-sided t-test) when we include all sociodemographic controls and baseline satisfaction with democracy.[17] When examining the two items separately, we find a clear negative treatment effect on the freedom of speech item (Table VI, Column 2), statistically significant at the 5% level when including all relevant controls. The media-independence analysis variable is also negative, but it is not statistically significant (Table VI, Column 3).[18] For the other items on democratic norms and trust in institutions, we find only small and statistically insignificant effects (Appendix K.2). Effects on the filler items are likewise insignificant, with a single exception (support for refugees, significant at the 5% level), which is consistent with chance given the number of outcomes tested (Appendix K.3).

Although we did not preregister an analysis by political orientation, the treatment effect may vary across the political spectrum: in many EU countries left-wing movements have historically maintained closer ties to Russia, while right-wing groups have more recently portrayed Russia as a political model. Differences in perceived threat, trust in EU institutions, and media consumption habits could therefore produce systematic variation in how participants respond to information about the media ban.

Figure VI shows the heterogeneous effect of the treatment by political stance. We split participants according to their self-reported political orientation, elicited at the end of the survey. Respondents could indicate that they lean left, center, or right, or choose not to report their orientation. The figure conveys a clear pattern: the treatment effect is concentrated among respondents who place themselves in the center of the political spectrum (N=182). Participants who identify as left-leaning (N=429) or right-

[17] As we preregister the direction of the expected effect, we also conduct a one-sided t-test, which is statistically significant at the 5% level. The magnitude of the effect corresponds to about 0.09 of a standard deviation in the freedom of speech index. Our sample was designed ex ante to detect a minimum effect size of 0.2 of a standard deviation in a two-sided test.

[18] The media-independence item was elicited in reverse wording and coded for analysis so that higher values indicate stronger perceived EU protection of media independence.

TABLE VI.
IMPACT OF THE INFORMATION TREATMENT ON FREEDOM OF SPEECH SATISFACTION

Dependent variable	**Freespeech index**	**Freedom of speech**	**Media independence**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Media ban information	-0.120 (0.072) [0.095]	-0.163 (0.079) [0.040]	-0.075 (0.100) [0.452]
Language FEs	✓	✓	✓
Baseline satisfaction with democracy	✓	✓	✓
Individual characteristics	✓	✓	✓
Mean dep. var.	0.484	0.932	0.031
Observations	800	803	800

Notes: The table reports the treatment effect of being exposed to a brief informing about the ban of Russia Today and Sputnik in our survey experiment (see Section 8 for details). Column 1 uses the pre-registered *Freespeech Index* as the dependent variable. It is computed as the average of the *Freedom of speech* and *Media Independence* analysis variables recorded separately in columns 2 and 3. All items were elicited on a 7-point Likert approval scale and coded/rescaled for analysis so that higher values indicate stronger perceived EU protection of the corresponding construct. The media-independence item was elicited in reverse wording and aligned to this higher-is-more-protection direction. We drop respondents who fail the attention check and have a duration to complete the survey below the 5th and above the 95th percentile in the duration distribution to exclude unreliable respondents. All specifications include language fixed effects and the full set of sociodemographic controls, as well as the baseline satisfaction with democracy elicited before the treatment was administered.

leaning (N=87) do not show any negative response to the treatment.[19] If anything, the point estimate for right-leaning respondents goes in the opposite direction.

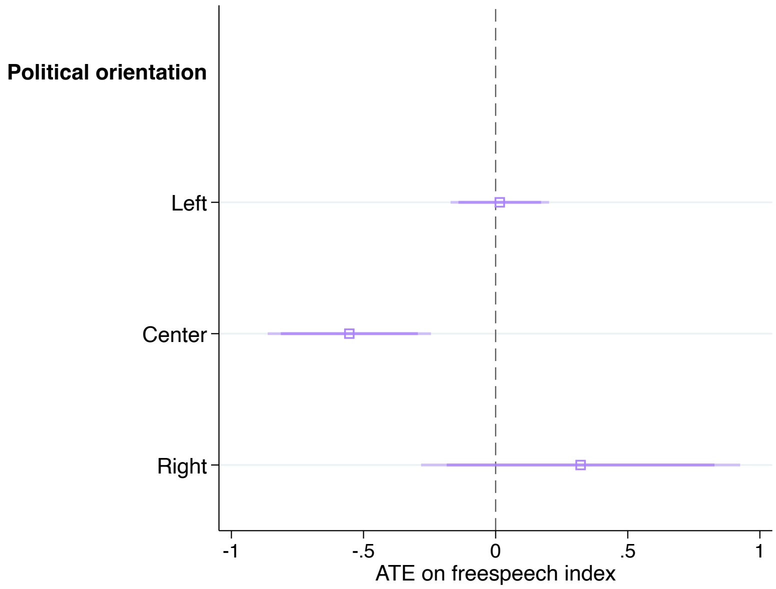
In sum, the experiment provides suggestive evidence of a cost of censorship as a counter-disinformation tool. The effect we recover is largely domain-specific – concentrated on perceptions of freedom of speech rather than broader measures of trust in the democratic order – and statistically modest. It is concentrated among political centrists, while respondents at the fringes of the political spectrum and those already dissatisfied with democracy do not show a comparable negative response. We also return to the open question of whether this pattern reflects rational updating about EU free-speech protections or a shift in underlying trust – a distinction the design cannot resolve.

9 Conclusion

In this paper, we study the effects of using censorship in a democratic context. We use the European Council's 2022 decision to ban all broadcasting activity of Russia Today and Sputnik to provide evidence on how the ban affected the production of pro-Russia content in a war-related Twitter corpus. A triple-difference design that exploits cross-user variation in pre-ban connection to the banned outlets recovers a

[19] We also find that the treatment effect is concentrated among respondents with high baseline satisfaction with democracy in the EU (N=448), whereas those with low satisfaction (N=352) do not appear to respond to the treatment (see Figure K.2).

FIGURE VI.
EXPERIMENT: HETEROGENEITY



Notes: The figure shows the treatment effect of being exposed to a brief informing about the ban of Russia Today and Sputnik in our survey experiment (see Section 8 for details) by the political orientation of the respondent. Dependent variable is the *Freespeech Index* computed as the average of the *Freedom of speech* and *Media Independence* analysis variables. All items were elicited on a 7-point Likert approval scale and coded/rescaled for analysis so that higher values indicate stronger perceived EU protection of the corresponding construct. The media-independence item was elicited in reverse wording and aligned to this higher-is-more-protection direction. We drop respondents who fail the attention check and have a duration to complete the survey below the 5th and above the 95th percentile in the duration distribution to exclude unreliable respondents. All specifications include language fixed effects and the full set of sociodemographic controls, as well as the baseline satisfaction with democracy elicited before the treatment was administered.

21.7% reduction in pro-Russia output among connected EU users, where connectedness proxies exposure to the outlets' content flow. A complementary two-way difference-in-differences shows a 13.6% reduction in the sampled corpus, of which the overwhelming majority appears in residual posting rather than direct outlet-linked content: the ban is associated with a shift in what users post, not only with the small share of content directly citing the outlets. We find no evidence of substitution by alternative pro-Russia suppliers within the sample – pre-ban active producers reduce their pro-Russia output rather than ramping it up to fill the gap – and no evidence of content-specific audience disengagement. The mechanism interpretation we emphasize is agenda-setting: the evidence is consistent with RT and Sputnik serving as real-time agenda-setters whose daily framings the connected user base tracked, and removing access to them broke that synchronization. The 17% post-ban drop in topical alignment between treated users and the outlets' daily top-5 topics is one empirical signature of this mechanism.

These findings come with several caveats that bound their interpretation. First, we observe only a narrow time window around the ban and cannot assess longer-term dynamics. Persistence, adaptation, and institutional or behavioral substitution that unfold over months rather than weeks fall outside our two-week post-ban window. Second, we observe activity on only one platform, Twitter, and only within a keyword-restricted war-related corpus. We cannot rule out substitution outside the sampled corpus or platform. Cross-platform substitution patterns have been documented in other settings (Rizzi 2024), though we view this as less of a concern here because the EU measure was not platform-specific but applied to all broadcasting activity of Russia Today and Sputnik. Third, our findings come from a single ban event in a specific high-stakes geopolitical context – the early phase of the Russian-Ukrainian war. Generalizing the patterns we document to bans on foreign state media adopted under different conditions, or to other restrictions on speech, requires additional evidence from other policy episodes. Fourth, the survey experiment in Section 8 recovers a modest effect, and the considerable gap between the implementation of the ban in 2022 and the survey fielding in 2025 limits how directly it speaks to the immediate aftermath of the policy. The design also cannot distinguish whether the treatment changes attitudes by providing new information about the ban or by raising the salience of an already-known policy.

Together, these findings point to a latent tension at the heart of using censorship as a tool of democratic self-defense. The ban reduced pro-Russia content within the observed Twitter corpus, primarily through shifts in residual posting rather than through the direct removal of outlet-linked content. The survey experiment, more modest and suggestive, points to a possible cost: making the ban salient lowers perceived EU commitment to freedom of speech and media independence among political centrists. The pattern speaks to the *Paradox of Tolerance* (Popper 1945): democracies may need restrictive measures to counter the threat of foreign propaganda, but those measures may carry a cost to the norms of free speech and media independence that define the democratic order they are meant to defend. Our paper offers strong observational evidence on the short-run content response and complementary suggestive evidence on democratic costs.

References

Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun (2018). "The Power of the Street: Evidence from Egypt's Arab Spring". In: *The Review of Financial Studies* 31.1, pp. 1–42.

Adena, Maja et al. (2015). "Radio and the Rise of The Nazis in Prewar Germany". In: *The Quarterly Journal of Economics* 130.4, pp. 1885–1939.

Allcott, Hunt and Matthew Gentzkow (2017). "Social Media and Fake News in the 2016 Election". In: *Journal of Economic Perspectives* 31.2, pp. 211–236.

Allcott, Hunt et al. (2020). "The Welfare Effects of Social Media". In: *American Economic Review* 110.3, pp. 629–676.

Andres, Raphaela and Olga Slivko (2021). "Combating Online Hate Speech: The Impact of Legislation on Twitter". In: *No. 21-103. ZEW Discussion Papers.*

Aridor, Guy et al. (Dec. 1, 2024). "The Economics of Social Media". In: *Journal of Economic Literature* 62.4, pp. 1422–1474.

Baker, Andrew et al. (June 1, 2026). "Difference-in-Differences Designs: A Practitioner's Guide". In: *Journal of Economic Literature* 64.2, pp. 498–557.

Barbera, Pablo (2014). "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the US". In: *Job Market Paper, New York University* 46, pp. 1–46.

Becker, Sascha O, Francisco J Pino, and Jordi Vidal-Robert (2021). "Freedom of the Press? Catholic Censorship During the Counter-Reformation". In: *CEPR Discussion Paper No. DP16092.*

Bjørnskov, Christian and Stefan Voigt (2021). "Is Constitutionalized Media Freedom Only Window Dressing? Evidence from Terrorist Attacks". In: *Public Choice* 187.3, pp. 321–348.

Blasutto, Fabio and David de la Croix (2023). "Catholic Censorship and the Demise of Knowledge Production in Early Modern Italy". In: *The Economic Journal* 133.656, pp. 2899–2924.

Bubeck, Sébastien et al. (2023). "Sparks of Artificial General Intelligence: Early Experiments with GPT-4". In: *arXiv preprint arXiv:2303.12712*. Version Number: 5.

Bursztyn, Leonardo et al. (2019). "Social Media and Xenophobia: Evidence from Russia". In: *No. w26567. National Bureau of Economic Research.*

Cage, Julia, Nicolas Herve, and Beatrice Mazoyer (2022). "Social Media and Newsroom Production Decisions". In: *SSRN Electronic Journal.*

Campante, Filipe, Ruben Durante, and Francesco Sobbrio (2018). "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation". In: *Journal of the European Economic Association* 16.4, pp. 1094–1136.

Campante, Filipe, Ruben Durante, and Andrea Tesei, eds. (2023). *The Political Economy of Social Media.* Paris & London: CEPR Press. 211 pp.

Chen, Jiafeng and Jonathan Roth (2024). "Logs with Zeros? Some Problems and Solutions". In: *The Quarterly Journal of Economics* 139.2, pp. 891–936.

Chen, Yuyu and David Y. Yang (2019). "The Impact of Media Censorship: 1984 or Brave New World?" In: *American Economic Review* 109.6, pp. 2294–2332.

Chu, Zi et al. (2012). "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" In: *IEEE Transactions on Dependable and Secure Computing* 9.6, pp. 811–824.

Corduneanu-Huci, Cristina and Alexander Hamilton (2022). "Selective Control: The Political Economy of Censorship". In: *Political Communication* 39.4, pp. 517–538.

Durante, Ruben and Brian Knight (June 2012). "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy". In: *Journal of the European Economic Association* 10.3, pp. 451–481.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei (July 1, 2019). "The Political Legacy of Entertainment TV". In: *American Economic Review* 109.7, pp. 2497–2530.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya (2011). "Media and Political Persuasion: Evidence from Russia". In: *American Economic Review* 101.7, pp. 3253–3285.

Gehlbach, Scott and Konstantin Sonin (2014). "Government Control of the Media". In: *Journal of Public Economics* 118, pp. 163–171.

Gehring, Kai and Matteo Grigoletto (2023). "Analyzing Climate Change Policy Narratives with the Character-Role Narrative Framework". In: *SSRN Electronic Journal*.

– (2025). "Virality: What Makes Narratives Go Viral, and Does it Matter". In: *Working Paper*.

Gennaro, Gloria and Elliott Ash (2023). "Emotion and Reason in Political Language". In: *The Economic Journal* 133.650, pp. 904–904.

Gentzkow, Matthew and Jesse M. Shapiro (2011). "Ideological Segregation Online and Offline". In: *The Quarterly Journal of Economics* 126.4, pp. 1799–1839.

Gentzkow, Matthew and Jesse M. Shapiro (Apr. 2006). "Media Bias and Reputation". In: *Journal of Political Economy* 114.2, pp. 280–316.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya (2021). "3G Internet and Confidence in Government". In: *The Quarterly Journal of Economics* 136.4, pp. 2533–2613.

Guriev, Sergei and Daniel Treisman (2022). *Spin Dictators: The Changing Face of Tyranny in the 21st Century*. 2023rd ed. Princeton University Press.

Guriev, Sergei et al. (2023). "Curtailing False News, Amplifying Truth". In: *SSRN Electronic Journal*.

Halberstam, Yosh and Brian Knight (2016). "Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter". In: *Journal of Public Economics* 143, pp. 73–88.

Jiménez-Durán, Rafael (2023). "The Economics of Content Moderation: Theory and Experimental Evidence from Hate Speech on Twitter". In: *George J. Stigler Center for the Study of the Economy & the State Working Paper 324*.

Jiménez Durán, Rafael, Karsten Müller, and Carlo Schwarz (2025). "The Effect of Content Moderation on Online and Offline Hate: Evidence from Germany's NetzDG". In: *SSRN Electronic Journal*.

Journalists, European Federation of (Mar. 1, 2022). *Fighting Disinformation with Censorship is a Mistake*. EFJ.

Kellam, Marisa and Elizabeth A. Stein (2016). "Silencing Critics: Why and How Presidents Restrict Media Freedom in Democracies". In: *Comparative Political Studies* 49.1, pp. 36–77.

Kozyreva, Anastasia et al. (May 13, 2024). “Toolbox of Individual-Level Interventions Against Online Misinformation”. In: *Nature Human Behaviour* 8.6, pp. 1044–1052.

Lazer, David M. J. et al. (Mar. 9, 2018). “The Science of Fake News”. In: *Science* 359.6380, pp. 1094–1096.

Levy, Ro’ee (2021). “Social Media, News Consumption, and Polarization: Evidence from a Field Experiment”. In: *American Economic Review* 111.3, pp. 831–870.

Linz, Juan J. (1978). *The Breakdown of Democratic Regimes.* Johns Hopkins University Press.

Liu, Pengfei et al. (2023). “Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing”. In: *ACM computing surveys* 55.9. Version Number: 1, pp. 1–35.

Morales, Juan S. (2020). “Perceived Popularity and Online Political Dissent: Evidence from Twitter in Venezuela”. In: *The International Journal of Press/Politics* 25.1, pp. 5–27.

Mullainathan, Sendhil and Andrei Shleifer (2005). “The Market for News”. In: *American Economic Review* 95.4, pp. 1031–1053.

Müller, Karsten and Carlo Schwarz (2022). “The Effects of Online Content Moderation: Evidence from President Trump’s Account Deletion”. In: *Available at SSRN 4296306.*

– (2023). “From Hashtag to Hate Crime: Twitter and Antiminority Sentiment”. In: *American Economic Journal: Applied Economics* 15.3, pp. 270–312.

Ni, Jianmo et al. (2021). “Sentence-T5: Scalable Sentence Encoders from Pre-trained Text-to-Text Models”. In: *arXiv preprint arXiv:2108.08877.* Version Number: 3.

OpenAI et al. (2023). *GPT-4 Technical Report.* Version Number: 6.

Pennycook, Gordon and David G. Rand (Apr. 28, 2022). “Accuracy Prompts Are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation”. In: *Nature Communications* 13.1, p. 2333.

Pennycook, Gordon et al. (July 2020). “Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention”. In: *Psychological Science* 31.7, pp. 770–780.

Persily, Nathaniel and Joshua A. Tucker, eds. (Aug. 31, 2020). *Social Media and Democracy: The State of the Field, Prospects for Reform.* 1st ed. Cambridge University Press.

Popović, Igor (Mar. 30, 2022). *The EU Ban of RT and Sputnik: Concerns Regarding Freedom of Expression.* Blog of the European Journal of International Law.

Popper, Karl R. (1945). *The Open Society and Its Enemies.* London: Routlege.

Rizzi, Marina (2024). “Self-Regulation of Social Media and the Evolution of Content: a Cross-Platform Analysis”. In: *Working Paper.*

Shadmehr, Mehdi and Dan Bernhardt (2015). “State Censorship”. In: *American Economic Journal: Microeconomics* 7.2, pp. 280–307.

Simonov, Andrey and Justin Rao (2022). “Demand for Online News Under Government Control: Evidence from Russia”. In: *Journal of Political Economy* 130.2, pp. 259–309.

Tabassum, Fatima et al. (2023). “How Many Features Do We Need to Identify Bots on Twitter?” In: *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity*, pp. 312–327.

Vériter, Sophie L (Sept. 5, 2025). “The Collective Securitization of ‘Disinformation’ and the EU’s Ban on *Russia Today* and *Sputnik*”. In: *International Affairs* 101.5, pp. 1853–1876.

Yanagizawa-Drott, David (2014). “Propaganda and Conflict: Evidence from the Rwandan Genocide”. In: *The Quarterly Journal of Economics* 129.4, pp. 1947–1994.

Young, Alwyn (May 1, 2019). “Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results”. In: *The Quarterly Journal of Economics* 134.2, pp. 557–598.

Zhao, Tony Z et al. (2021). “Calibrate Before Use: Improving Few-Shot Performance of Language Models”. In: *Proceedings of the 38th International Conference on Machine Learning* 139, pp. 12697–12706.

Censorship in Democracy

Marcel Caesmann*, Matteo Grigoletto**, Lorenz Gschwent°°

Online Appendix

*University of Zurich, e-mail: *marcel.caesmann@econ.uzh.ch*

**University of Bern, Wyss Academy for Nature at the University of Bern, e-mail: *matteo.grigoletto@unibe.ch*

°°University of Duisburg-Essen, e-mail: *lorenz.gschwent@uni-due.de*

Appendix

Table of Contents

A Materials and Methods

This section provides insights into the extraction, processing, and classification of Twitter data, expanding on the information provided in 4.1.

(i) **Query definition.** The download of tweets is done via the Twitter APIv2 that allows researchers to extract any tweets posted and not deleted in the platform since 2006, with a monthly cap of ten million tweets.[20] Our dataset consists of tweets posted by users involved in the discussion about the unfolding conflict and the later invasion of Ukraine. When extracting this data, a clear trade-off emerges. On one side, we want to ensure that we capture a representative sample of the conversation about the conflict. Hence, it is necessary to use a query that is not too restrictive. On the other side, we need to impose some restrictions to avoid false positives – tweets not primarily concerned with the conflict. Hence, our keyword query to solve this issue focuses on the main entities involved in the conflict: Russia, Ukraine, and NATO; this led to the following query: *russ* OR ukrain* OR nato OR otan*.

(ii) **Initial extraction.** Using the query above, we initially downloaded all tweets fulfilling these conditions posted between January 24$^{\text{th}}$, 2022, and April 4$^{\text{th}}$, 2022, and the following day-hours windows: 9 a.m. to 12 p.m., 3 p.m. to 6 p.m., and 8 p.m. to 11 p.m. This results in 7,865,321 extracted tweets by 1,942,979 users. This initial dataset includes many users with either missing geolocation information or users located outside of the policy setting we study.

(iii) **Geolocation.** To restrict the sample to users who allow analysis of the EU's ban, we employ a geolocation strategy. Utilizing the geolocation method outlined in Gehring and Grigoletto (2025), we use locations, for example, countries, states, or cities, that users have indicated in their Twitter profiles to assign a location to users. Then, we identify users located in our target countries: Austria, France, Germany, Ireland, Italy, Switzerland, and the UK. After restricting to users with at least one tweet in the 22-day analysis window between February 22$^{\text{nd}}$ and March 15$^{\text{th}}$, the analysis sample comprises 146,633 geolocated users in our target countries. This is the sample of users who are tweeting about the Russo-Ukrainian conflict that we use for our analysis.

(iv) **Final download.** As the initial sweep of tweets concerning the conflict in step (ii) only retrieves tweets in specific time windows, we next download all tweets matching our query (*russ* OR ukrain* OR nato OR otan*) by the 146,633 users on which we settled via the previous steps. This process yields a dataset of 677,780 tweets (original tweets, replies, and retweets with no more than 140 characters, as discussed in the next subsection) posted between February 22$^{\text{nd}}$ and March 15$^{\text{th}}$.

(v) **Classification.** We classify our data using an OpenAI Large Language Model, accessed via the API. To streamline our classification pipeline, we only use tweets for which we were able to retrieve the full text via the Twitter API. In 2017, Twitter increased the character limit from 140 to 280 characters per tweet. However, a retweet of an original tweet longer than 140 characters is truncated

[20] For more information, see the Twitter Developer Platform documentation of the Search Tweets endpoint.

when we retrieve it via the API. To avoid any misclassifications caused by the truncated text, we only use untruncated retweets in our classification and further analysis. The classification process, then, is straightforward. Tweets are organized into batches and sent to the API for classification. OpenAI enforces a zero data retention policy for API usage: submitted data are not stored after processing and are never used to train or fine-tune the models. We rely on the model GPT-4o-mini, which offers a good compromise between accuracy and processing speed. A single prompt is applied uniformly to all tweets to ensure consistency and reliability across the full dataset. The exact prompt is reproduced below:

> *"You are an objective political analyst tasked with analyzing tweets related to the Russia-Ukraine war. Your goal is to classify whether a tweet contains pro-Russia content and, if so, identify key topics in the tweet. Respond strictly in JSON format.*
>
> *Context: The tweets you are analyzing were posted within a one-month window around the February 2022 full-scale invasion of Ukraine by Russia. The war has its roots in a broader geopolitical conflict that began in 2014 with Russia's annexation of Crimea and its involvement in the conflict in Eastern Ukraine. Your analysis should be conducted with this historical and political context in mind. Pro-Russia content refers to messaging that supports, justifies, or aligns with Russian interests in the context of this war.*
>
> *1. Primary Classification (pr): Assess whether the tweet explicitly contains pro-Russia content based on clear textual evidence (not intuition or inferred sentiment). Assign one of the following values: - 1 = Pro-Russia (The tweet aligns with Russian narratives, either overtly or through framing, or supports Russia in the conflict.) - 2 = Against Russia (The tweet opposes Russian narratives, criticizes Russia's actions, or supports Ukraine in the conflict.) - 0 = Neutral / Not Clear (The tweet does not explicitly take a position, is purely factual, or lacks sufficient information to determine alignment.)*
>
> *2. Topic Identification (t): Extract up to 5 topics present in the tweet. Each topic should be labeled using 1–3 words. If multiple topics are present, separate them using a semicolon (";"). Example topics: "NATO expansion"; "Ukraine military"; "war crimes"; "media bias"; "Western aid".*
>
> *3. JSON Output Format: { "id": "Unique identifier for the tweet.", "pr": 0, 1, or 2 (based on step 1), "t": "String containing up to 5 topics separated by ';' (based on step 2)" }"*

In addition to this main sample of Twitter users discussing the Russo-Ukrainian conflict, we also collect all tweets posted between January 24$^{\text{th}}$, 2022, and April 4$^{\text{th}}$, 2022 by Russia Today, Sputnik, the Russian news agency TASS (used in Subsection 6.5), as well as by Ukrainian and Russian government-associated accounts. For this last group, we provide a full list of the accounts in Table C.1 in Appendix C, where we describe an alternative text classification method relying on the tweets by these government accounts. The only filtering we do for this sample is on the language. To be part of this sample, a tweet has to come from the selected accounts, in the period of interest, and has to be in English.

B Additional Descriptive Output

In this Appendix, we present additional descriptive output, complementing what we show in the main paper. We start with Table B.1. It reports a list of all the variables used in the analysis. The variables are divided into *Outcomes*, for variables that are used as dependent variables, and *Factors*, for variables used to sub-sample the dataset for the analysis or as control variables. For each variable we provide a short description, an indication of its scale or range, and a short indication of the source of the variable.

Table B.2 lists the twenty accounts with the highest closeness centrality in the full sample, the EU, and non-EU pre-ban networks. Closeness measures an account's global position for information diffusion: it is the reciprocal of the average number of retweet-and-reply steps separating the account from all users who can reach it, so that a value of 0.30 implies the average user is roughly three steps removed from the focal account. Unlike in-degree, which counts an account's direct audience, closeness rewards positions from which content can reach the entire network through short chains. Because shortest-path distances are tightly concentrated in small-world networks, the measure is compressed, and ranks are the informative statistic. In the full sample, in a network of 80,404 connected users, RT/Sputnik ranks 30$^{\text{th}}$ (with a raw closeness centrality measure of 0.0205), as we show in the main paper's Figure I. In the EU network, RT/Sputnik ranks seventh out of 30,672 accounts forming the connected network, in the immediate vicinity of major European news outlets or the Twitter accounts of heads of state. In the non-EU network, by contrast, RT/Sputnik ranks 36$^{\text{th}}$ (with a raw closeness centrality measure of 0.3015) out of 17,678 accounts forming the connected network.

TABLE B.1.
DESCRIPTION OF THE VARIABLES USED IN THE ANALYSIS

Variable	Description	Scale/Range	Source
Outcomes			
Pro-Russia Content	Indicator variables capturing whether a tweet (”own tweets”, replies, or retweets with no more than 140 characters) features pro-Russia content	$n \in [0;1]$	OpenAI API: GPT-4o
Absolute No. of pro-Russia tweets	Number of tweets (”own tweets”, replies, or retweets with no more than 140 characters) containing pro-Russia content by unit of observation, conditional on posting about the war	$n \in [0;263]$	Own computation
Likelihood of posting pro-Russia	Indicator variable capturing whether the user posted pro-Russia content (”own tweets”, replies, or retweets with no more than 140 characters) in a given day, conditional on posting	$n \in [0;1]$	Own computation
Likelihood of posting about the war	Indicator variable capturing whether the user posted (”own tweets”, replies, or retweets with no more than 140 characters) at all about the war on a given day	$n \in [0;1]$	Own computation
Measure of engagement	Sum of retweets, replies, and likes, obtained by a tweet (”own tweets” or replies) featuring pro-Russia content	$n \in [0;27,530]$	Own computation
Topic rate	Measures the number of topics per pro-Russia tweet (”own tweets”, replies, or retweets with no more than 140 characters)	$n \in [0;6]$	OpenAI API: GPT-4o
# top 5 outlets' topics per tweet	Measures the overlap between the topics in a pro-Russia tweet (”own tweets”, replies, or retweets with no more than 140 characters) and the top 5 most common topics in the tweets of Russia Today and Sputnik on the same day	$n \in [0;4]$	OpenAI API: GPT-4o
Factors			
Connected	Indicator variable capturing whether the user is part of the network of Russia Today and Sputnik before the ban	$n \in [0;1]$	Own computation
Distance to the outlets	Degrees of distance between a user and Russia Today or Sputnik; the variable takes the value 0 for users not connected to the outlets; for connected users, the variable takes the value 1 for users who directly interacted with the outlets, 2 for users who interacted with users in group 1, and so on	$n \in [0;8]$	Own computation
Activity pre-ban	Days (out of 8) a user posted pro-Russia content (”own tweets” or replies) before the ban	$n \in [0;8]$	Own computation
High pre-ban activity	Indicator variable capturing whether a user posted pro-Russia content (”own tweets” or replies) on at least four days before the ban	$n \in [0;1]$	Own computation
Top engagement	Indicator variable capturing whether a user was in the top 5% of engagement (sum of retweets, replies, and likes) with their original pro-Russia content (”own tweets” or replies) before the ban	$n \in [0;1]$	Own computation

Notes: The table describes the variables used in the analysis, either as outcomes or as factors to sub-sample our dataset. For each variable, we provide a short description, the scale or range of the variable, and the source.

TABLE B.2.
CLOSENESS CENTRALITY RANKS

Panel A: Top 20 Full Network Handles

rank	username	value
1	BorisJohnson	0.1011
2	Telegraph	0.0507
3	SkyNews	0.0473
4	ZemmourEric	0.0451
5	SamRamani2	0.0347
6	OlafScholz	0.0334
7	DiegoFusaro	0.0334
8	mikegalsworthy	0.0326
9	JLMelenchon	0.0321
10	paul	0.0308
11	BBCRosAtkins	0.0296
12	Cartabellotta	0.0290
13	UPR_Asselineau	0.0284
14	L_ThinkTank	0.0273
15	bru	0.0262
16	FerghaneA	0.0260
17	Samfr	0.0253
18	DefenceHQ	0.0251
19	briceculturier	0.0244
20	nicktolhurst	0.0241

Panel B: Top 20 EU Handles

rank	username	value
1	ZemmourEric	0.3043
2	Nico	0.3037
3	JLMelenchon	0.3021
4	UPR_Asselineau	0.3003
5	LesEchos	0.2992
6	OlafScholz	0.2984
7	RT/Sputnik	0.2983
8	enne_lory	0.2978
9	MXThiel	0.2962
10	Chaosjana1980	0.2958
11	Guillaume__Lec	0.2953
12	RenStelzer1	0.2936
13	frarizz61	0.2925
14	AFP	0.2923
15	Pierre_GTIL	0.2915
16	L_ThinkTank	0.2915
17	Cecilia3196	0.2905
18	Polocartes	0.2895
19	afpfr	0.2893
20	derspiegel	0.2889

Panel C: Top 20 Non-EU Handles

rank	username	value
1	BorisJohnson	0.3565
2	Telegraph	0.3277
3	paul	0.3267
4	WriterJackWhite	0.3166
5	Andrew3745	0.3151
6	F15JCM	0.3150
7	WildlifeBloke	0.3142
8	SkyNews	0.3140
9	nicktolhurst	0.3126
10	MarcReporting	0.3125
11	SamRamani2	0.3116
12	SimonPease1	0.3111
13	arhmad_didi	0.3107
14	JohnPeach195	0.3107
15	DroningonUK	0.3097
16	guv360	0.3085
17	RonnySiev	0.3085
18	scribblercat	0.3082
19	RhonddaBryant	0.3079
20	paulwaugh	0.3070

Notes: Closeness centrality is computed on the directed retweet-and-reply network using all tweets in the pre-ban period (February 16 - March 2, 2022). Closeness centrality captures how quickly a node can reach all others in the network. Higher values indicate shorter average path lengths to all reachable nodes. Networks are constructed separately for the full sample (Panel A), for users located in EU member states (Panel B), and for European non-EU countries (Panel C). Accounts are pooled across all languages. *RT/Sputnik* denotes the merged account combining Russia Today and Sputnik News.

Table B.3 provides descriptive statistics for the main sample of tweets used in the analysis, complementing the discussion in Section 4. Table B.4 presents descriptive statistics for the 146,633 users in the analysis sample. We provide information on the level of activity of these users. On average, each user produced or shared 4.6 tweets in our sample, and one tweet, which was classified as pro-Russia. It is important to mention that for the users of our sample, we collected all tweets mentioning our keywords in the one month around the ban, which allows us to compile a user-day panel dataset. The table also shows that 6.2% of users are connected to the banned outlets under our definition (Section 5.1). Additionally, roughly 39% of users are located in the comparison countries, the United Kingdom and Switzerland, while the others are in the treated countries, Austria, France, Germany, Ireland, and Italy.

TABLE B.3.
FEATURES OF THE TWEETS USED IN THE ANALYSIS

PANEL A: TWEETS POSTED BY THE OUTLETS

	Mean	**Median**	**St. Dev.**	**Min.**	**Max.**
Dependent variables					
Pro-Russia content	.18	0	.38	0	1
# topics in tweet	2.1	3	1.7	0	5
RT vs. Sputnik					
Share of tweets by Russia Today	.64	1	.48	0	1
Observations	4,362				

PANEL B: TWEETS POSTED BY STANDARD USERS

	Mean	**Median**	**St. Dev.**	**Min.**	**Max.**
Dependent variables					
Pro-Russia content	.22	0	.42	0	1
# topics in tweet	2.9	3	1.5	0	5
Connection to Russia Today and Sputnik					
Connected to RT/Sputnik (degree ≤ 2 or direct interactor)	.19	0	.4	0	1
Location					
Tweet is from Austria	.021	0	.14	0	1
Tweet is from France	.26	0	.44	0	1
Tweet is from Germany	.21	0	.4	0	1
Tweet is from Ireland	.026	0	.16	0	1
Tweet is from Italy	.12	0	.32	0	1
Tweet is from Switzerland	.026	0	.16	0	1
Tweet is from United Kingdom	.34	0	.47	0	1
Observations	677,780				

Notes: The table presents descriptive statistics for the datasets of tweets used in our analysis. Panel A describes tweets posted by Russia Today and Sputnik. This dataset comprises all content captured in our sample described in Section 4.1 posted between February 22$^{\text{nd}}$, 2022, and March 15$^{\text{th}}$, 2022. We collect these tweets using the Twitter API, retrieving all tweets posted by these accounts during the sampling period. Panel B describes tweets posted by standard platform users. This dataset comprises all content captured in our sample described in Section 4.1. We collect these tweets using the Twitter API, retrieving all tweets containing at least one of the keywords “NATO OTAN Russia Ukraine” over the entire sampling period. For each variable, we report the mean, median, standard deviation, minimum, and maximum values. Appendix Figure B.1 shows the distribution of tweets by language.

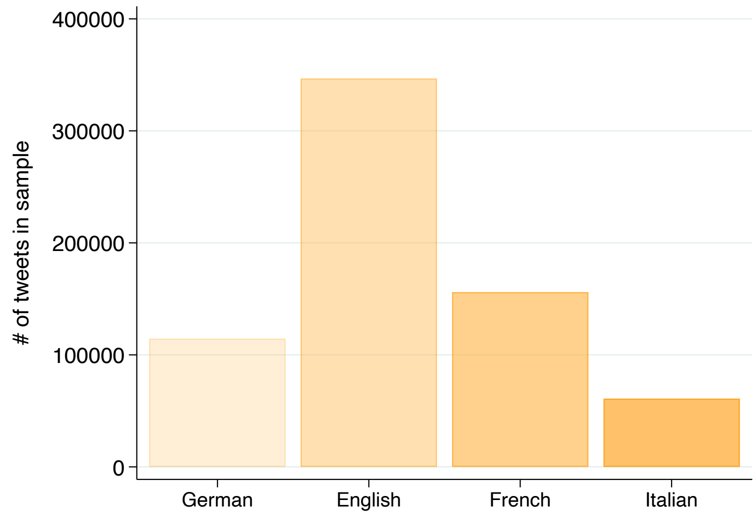
TABLE B.4.
USER LEVEL DESCRIPTIVES

	Mean	**Median**	**St. Dev.**	**Min.**	**Max.**
Activity					
Number of tweets	4.6	2	14	1	1,534
Number of pro-Russia tweets	1	0	4.6	0	827
Connection to Russia Today and Sputnik					
Connected to RT/Sputnik (degree $\leq$ 2 or direct interactor)	.062	0	.24	0	1
Location					
User is from Austria	.021	0	.14	0	1
User is from France	.23	0	.42	0	1
User is from Germany	.19	0	.39	0	1
User is from Ireland	.028	0	.16	0	1
User is from Italy	.14	0	.35	0	1
User is from Switzerland	.023	0	.15	0	1
User is from United Kingdom	.37	0	.48	0	1
No. of Observations	146,633				

Notes: The table presents descriptive statistics for the users in our analysis. This dataset comprises all content captured in our sample described in Section 4.1 posted between February 22$^{\text{nd}}$, 2022, and March 15$^{\text{th}}$, 2022. We collect these tweets using the Twitter API, retrieving all tweets containing at least one of the keywords "NATO OTAN Russia Ukraine" over the entire sampling period. For each variable, we report the mean, median, standard deviation, minimum, and maximum values.

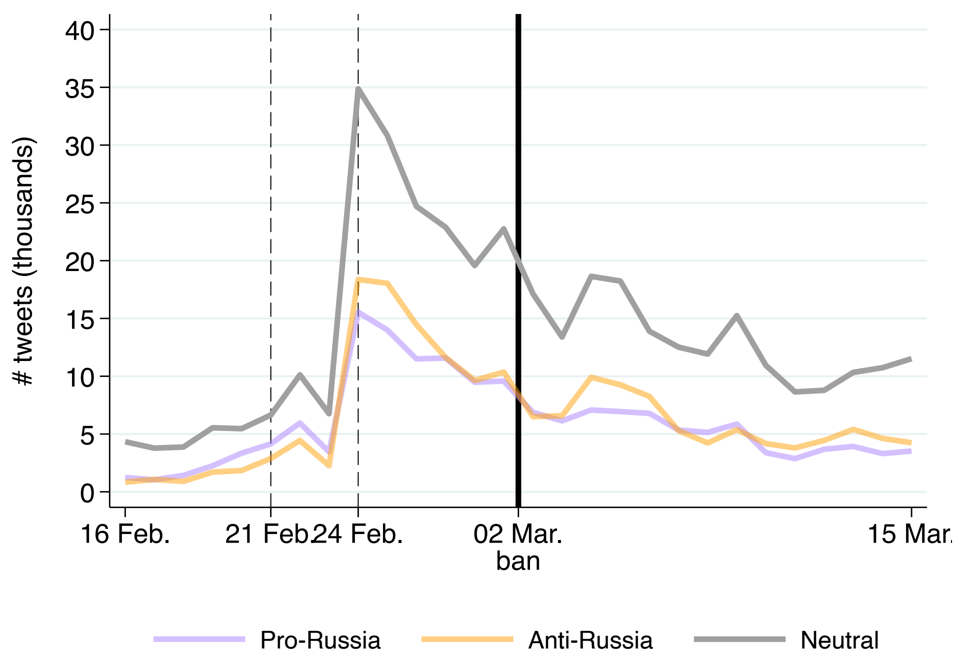
Figure B.1 complements the table above by providing a visualization of the number of tweets in our sample, divided by the language of the tweet. Although all tweets are translated into English before providing them as input to the Large Language Model used for classification, the tweets are collected as originally posted. We keep in our sample the four languages covered by our European sample: English, German, French, and Italian. As the figure shows, English language tweets account for over half of the corpus. This leads to a fairly balanced number of tweets between the control and treatment group, as most of the English tweets are coming from the United Kingdom.

FIGURE B.1.
TWEETS BY LANGUAGE



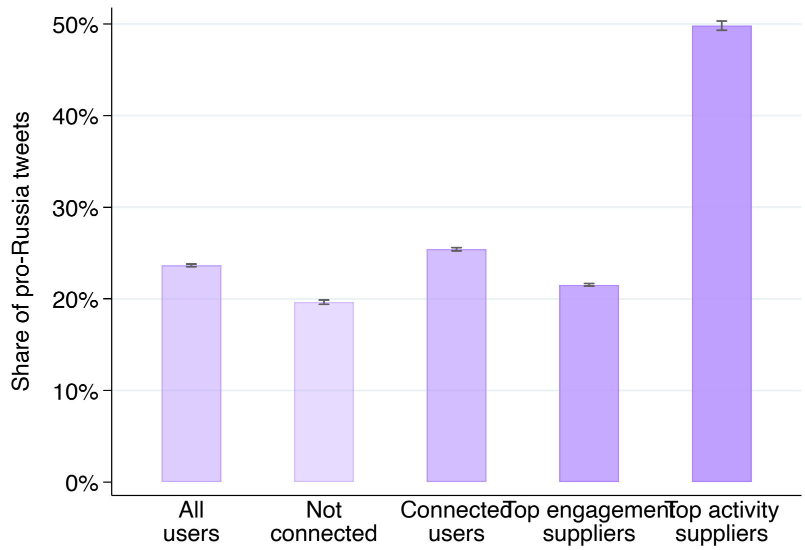
Notes: The figure shows the distribution of tweets in our sample by language of the original post. We translate all non-English tweets to English before putting them into the classification pipeline. The Table B.3 provides additional insights into the data used in the analysis.

FIGURE B.2.
TIME-SERIES OF THE MEASURES CLASSIFIED BY GPT



Notes: The figure displays the daily count of tweets classified by our pipeline containing pro-Russia content (in violet), anti-Russia content (in orange), and neutral content (in gray). The time series covers the period from February 16th, 2022, to March 15th, 2022, and pools together tweets originating from the EU countries (Austria, France, Germany, Ireland, and Italy) and the non-EU countries (United Kingdom and Switzerland). All other outputs of our analysis include only tweets posted on or after February 22nd, 2022, to ensure broad coverage and sufficient volume for statistical power in regression models. Vertical lines highlight key dates: February 21st, marking Putin's official recognition of the Donetsk People's Republic and the Luhansk People's Republic; February 24th, indicating the beginning of the war; and March 2nd, representing the start date of the ban.

FIGURE B.3.
SHARE OF PRO-RUSSIA ACTIVITY BEFORE THE BAN BY SUB-SAMPLE



Notes: The figure displays the share of tweets classified as pro-Russia content during the eight days preceding the ban, along with 95% confidence intervals, for each subsample of users included in our analysis. We include observations from February 22$^{\text{nd}}$ to March 1$^{\text{st}}$, 2022. The subsamples, shown from left to right in the figure, are as follows: (1) all users; (2) users with no direct interaction with the banned outlets; (3) direct interactors (users who replied to or retweeted the outlets at least once prior to the ban); (4) top-engagement pro-Russia producers (the top 5% of producers by engagement on their pro-Russia tweets); and (5) secondary suppliers (producers who posted pro-Russia content on at least four of the eight days before the ban). The first-degree direct-interactor labels in this figure are distinct from the degree-two connected indicator used in the main regressions.

Table B.5 sheds light on the topics that Russia Today and Sputnik chose to emphasize in their tweets. Recall that our LLM classifier performed two operations on every tweet in the corpus: (i) classified the tweets as pro-Russia, anti-Russia, or neutral in the context of the war, and (ii) tagged up to five topics featured in the tweet. The table restricts attention to tweets originating from the two outlets and, for each day in the observation window, lists the five most frequently tagged topics. Daily frequencies are calculated by summing the number of times a given topic appears anywhere in the outlet tweets posted on that date.

Several patterns stand out. Approaching Moscow's formal recognition of the self-proclaimed Luhansk and Donetsk republics, the content of the outlets' tweets focuses on *diplomacy* and the two regions themselves, consistent with an effort to frame the conflict as a matter of legitimate political choices. On 21$^{\text{st}}$ February, the recognition day, the topic *recognition* enters the top five topics list, where it remains the following day. Beginning 24$^{\text{th}}$ February, when the full-scale invasion starts, the narrative shifts toward *military operation*, echoing the Kremlin's preferred narrative about the war. The topic *sanctions* is a sustained focus across the analysis window, frequently among the top topics through the run-up to and aftermath of the ban, signaling an ongoing attempt to shape European opinion on the costs of restrictive measures. Taken together, the table shows how the outlets continuously re-aligned their messaging with major diplomatic, military, and policy milestones.

TABLE B.5.
TOP 5 OUTLET TOPICS BY DAY

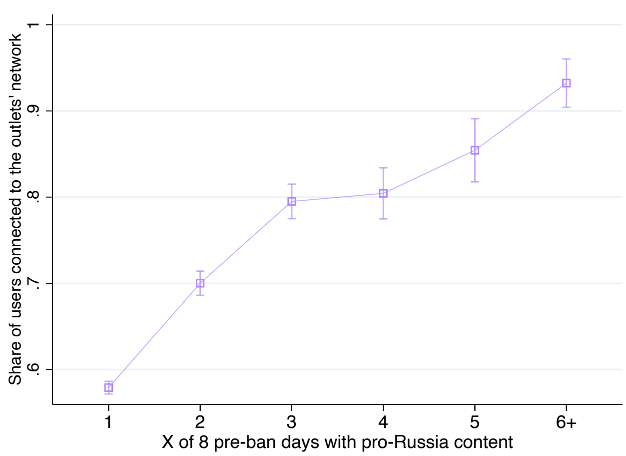
Date	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
2022-02-16	media bias	syria	belarus	china	military drills
2022-02-17	nato expansion	military drills	security guarantees	us	tensions
2022-02-18	donetsk	evacuation	donbass	media bias	explosion
2022-02-19	donetsk	nato expansion	evacuation	donbas	media
2022-02-20	donbass	diplomacy	donetsk	macron	shelling
2022-02-21	donbass	donetsk	recognition	nato expansion	lugansk
2022-02-22	sanctions	donbass	recognition	donetsk	lpr
2022-02-23	sanctions	donbass	dpr	nord stream 2	media bias
2022-02-24	donbass	military operation	sanctions	lugansk	special operation
2022-02-25	sanctions	media bias	military operation	zelensky	diplomacy
2022-02-26	sanctions	media	media bias	swift	germany
2022-02-27	media bias	sanctions	eu sanctions	belarus	china
2022-02-28	sanctions	media bias	negotiations	belarus	europe
2022-03-01	media bias	sanctions	conflict	un	europe
2022-03-02	sanctions	media bias	diplomacy	censorship	military operation
2022-03-03	sanctions	media	media bias	humanitarian aid	rt
2022-03-04	sanctions	nato expansion	media bias	nuclear power	military
2022-03-05	media bias	media	sanctions	mariupol	nuclear power
2022-03-06	donbass	media bias	conflict	sanctions	zelensky
2022-03-07	media bias	diplomacy	talks	sanctions	china
2022-03-08	sanctions	military operation	media bias	china	russian oil
2022-03-09	sanctions	donbass	media bias	us	civilians
2022-03-10	sanctions	diplomacy	turkey	media bias	mariupol
2022-03-11	media bias	sanctions	violence	meta	zelensky
2022-03-12	sanctions	media bias	violence	russian forces	russian military
2022-03-13	sanctions	media bias	china	conflict	us sanctions
2022-03-14	sanctions	media bias	china	donetsk	moscow
2022-03-15	sanctions	china	civilians	russian military	mariupol

Notes: The table reports the top 5 most used topics in the tweets posted by the banned outlets: Russia Today and Sputnik. We obtain the list of topics through our LLM classification pipeline. We queried the model to provide a list of up to five topics features in each tweet that we put into the pipeline. The Paper Table V is based on these topics.

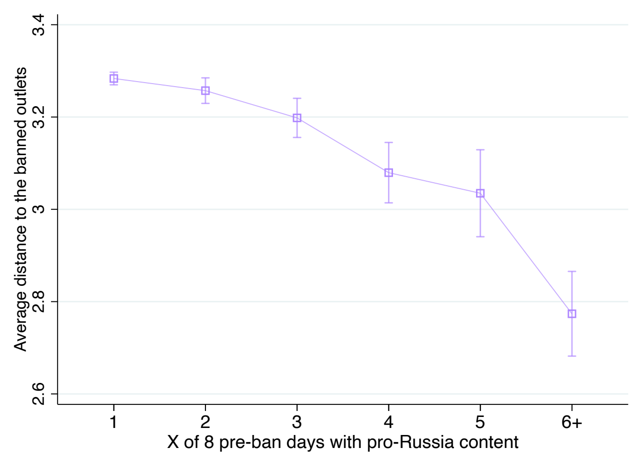
Figures B.4 and B.5 show the connection between the pre-ban activity of pro-Russia producers and other characteristics: being part of the networks of Russia Today and Sputnik, and the engagement that producers' pro-Russia content managed to attract before the ban. We document the co-movement of activity, network membership, distance to the banned outlets within the network, and engagement. As all of these variables are potentially relevant dimensions when it comes to differences in the ban's effect, we address the overlap between producer characteristics by forming mutually exclusive combinations of user groups for our heterogeneity analysis. The figures span the full activity gradient (1 to 6+ pre-ban days posting pro-Russia content); the secondary-supplier subset (the high-activity bin) is where outlet connection, network proximity, and engagement all concentrate.

FIGURE B.4.
PRO-RUSSIA PRODUCERS' ROLES IN THE OUTLETS' NETWORK BY LEVEL OF PRE-BAN ACTIVITY

(a) PRO-RUSSIA PRODUCERS' CONNECTIONS TO RT AND SPUTNIK

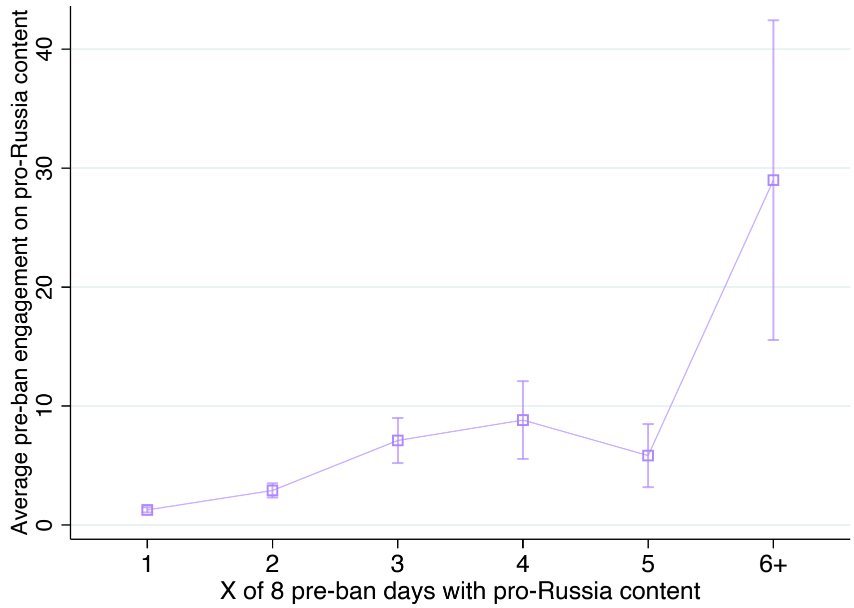


(b) PRO-RUSSIA PRODUCERS' DISTANCE TO RT AND SPUTNIK



Notes: The figure visualizes correlations of pro-Russia producers' days with pro-Russian activity before the ban and other user characteristics. In Panel A, we show the average share of users being part of the network of Russia Today and Sputnik before the ban. In Panel B, we plot the average distance to Russia Today or Sputnik for users who were part of the network before the ban. Users have a distance equal to 1 if they directly interacted (retweeted or replied to) the outlets before the ban. The distance value is equal to 2 for users who directly interacted with users in group 1, and so on. The descriptive "connected" indicator in Panel A uses any network degree to RT/Sputnik; the analytical specifications in Section 5.1 and the appendix tables use the tighter definition (degree ≤ 2 or direct interactor) introduced in Section 4.1.

FIGURE B.5.
PRO-RUSSIA PRODUCERS' PRE-BAN ENGAGEMENT BY LEVEL OF PRE-BAN ACTIVITY



Notes: The figure plots the average pre-ban engagement of pro-Russia tweets relative to the number of pre-ban days with pro-Russia activity. Sample: pro-Russia producers (any user with at least one pre-ban day posting pro-Russia content); the analytical sample in Table IV restricts to the secondary-supplier subset (active on ≥ 4 of 8 pre-ban days).

C Validation of GPT-Coded Pro-Russia Measurement

There is growing evidence that AI models built on large language models (LLM) perform reliably across many natural-language-processing and text-analysis tasks (Bubeck et al. 2023; OpenAI et al. 2023). In this paper we classify our corpus with OpenAI's GPT-4o-mini, with a pipeline introduced in Appendix A, to obtain a label of pro-Russia, neutral, or anti-Russia content for each tweet. Although recent evaluations confirm the high performance of GPT-4 variants (Bubeck et al. 2023), this appendix validates our coding by benchmarking GPT-4o-mini against a different way to measure pro-Russia content.

For this alternative measure, we conceptualize the discourse on the war as defined by a one-dimensional continuum between two narrative poles: pro-Russia and pro-Ukraine. In our analysis, discussions about the conflict occupy positions along this spectrum, with a tweet's content being closer to one pole or the other, indicating its narrative slant. This proximity to either pole reflects the intensity of its alignment, with content equidistant from both poles representing a neutral stance, thus supporting neither side strongly.

To obtain a quantifiable and tractable measure of pro-Russian and pro-Ukrainian media slant, we adopt a procedure proposed by Gennaro and Ash (2023), drawing inspiration from earlier work on media slant by Gentzkow and Shapiro (2011). This approach is both simple and powerful, relying on the measurement of language similarity in tweets by European users discussing the war relative to two distinct ideological poles. More specifically, we calculate the cosine similarity of a tweet's language to what we define as the 'pro-Russian pole' and compare this to its similarity to what we define as the 'pro-Ukrainian pole'. The critical decision in this method lies in determining the content that constitutes each pole.

We construct our ideological poles using content disseminated on Twitter by key figures within the Russian and Ukrainian governments. To achieve this, we systematically gather tweets posted by accounts affiliated with these governments, the comprehensive list of which is detailed in Table C.1. Our collection encompasses 5,993 tweets from Russian government representatives and 9,451 tweets from Ukrainian government representatives, collected over the period between January 24$^{\text{th}}$, 2022, to April 4$^{\text{th}}$. This dataset proves instrumental in establishing our measure of media slant, offering a direct insight into the narratives each government endorsed and ensuring our analytical framework's robustness.

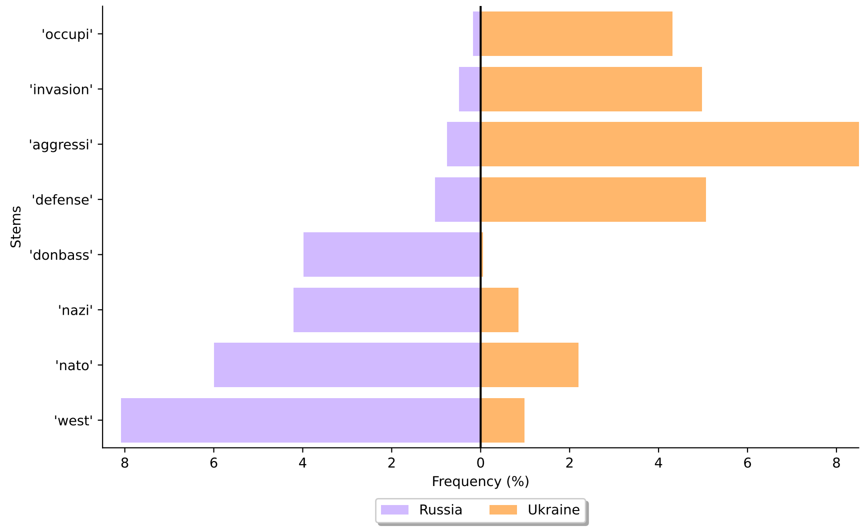
TABLE C.1. ACCOUNTS OF THE RUSSIAN AND UKRAINIAN GOVERNMENTS' REPRESENTATIVES

Ukrainian Accounts	Account Holder	Russian Accounts	Account Holder
https://twitter.com/DI_Ukraine	Defence Intelligence	https://twitter.com/RussianEmbassy	Embassy in the UK
https://twitter.com/Ukraine	Ukraine	https://twitter.com/mfa_russia	Ministry of Foreign Affairs
https://twitter.com/DefenceU	Ministry of Defense	https://twitter.com/mission_rf	Mission to the International Organizations in Vienna
https://twitter.com/CinC_AFU	Commander in Chief of Ukrainian Armed Forces	https://twitter.com/RF_OSCE	Mission to the OSCE
https://twitter.com/oleksiireznikov	Minister of Defence	https://twitter.com/RusEmbUSA	Embassy in the US
https://twitter.com/kabmin_ua_e	Cabinet of Ministers	https://twitter.com/RussianEmbassyC	Embassy in Canada
https://twitter.com/MFA_Ukraine	Ministry of Foreign Affairs	https://twitter.com/KremlinRussia_E	Official Kremlin News
https://twitter.com/DmytroKuleba	Minister of Foreign Affairs	https://twitter.com/EmbassyofRussia	Embassy in South Africa
https://twitter.com/AndriyYermak	Head of the Office of the President	https://twitter.com/PMSimferopol	Ministry of Foreign Affairs' Office in Crimea
https://twitter.com/NSDC_ua	Press Service of the National Security and Defense Council	https://twitter.com/RusMission_EU	Mission to the EU
https://twitter.com/UKRinDEU	Embassy of Ukraine in Germany	https://twitter.com/RusBotschaft	Embassy in Germany
https://twitter.com/ukrinche	Embassy of Ukraine in Switzerland	https://twitter.com/RusEmbSwiss	Embassy in Switzerland
https://twitter.com/ukrinfra	Embassy of Ukraine in France	https://twitter.com/ambrusfrance	Embassy in France
https://twitter.com/ukrinit	Embassy of Ukraine in Italy	https://twitter.com/rusembitaly	Embassy in Italy
https://twitter.com/UkrEmbLondon	Embassy of Ukraine in the UK		
https://twitter.com/MelnykAndrij	Ukrainian Ambassador to Germany		

Notes: The table reports the Russian and Ukrainian government-affiliated accounts that were used as sources for the two poles used to create our slant measurement. For each of these accounts, we extracted tweets in English between January 24$^{\text{th}}$, 2022, and April 4$^{\text{th}}$, 2022. This extraction resulted in 5,993 tweets for the Russian pole and 9,451 tweets for the Ukrainian pole. Note that to increase the number of English accounts on the Russian side, we also included the embassy account for non-European English-speaking countries active on Twitter.

Figure C.1 provides insights into the content of these government tweets through keyword frequency analysis. We differentiate the frequencies of Russian and Ukrainian government tweets, represented in violet and orange, respectively. Keywords like 'aggression' and 'invasion' are predominantly used by Ukrainian accounts to portray the conflict as an invasion, contrasting with the Russian portrayal as a 'military operation'. Other stems like 'occupi', 'defense', 'nato', 'west', 'nazi', and 'donbass' further delineate the narratives of each side. The use of these terms underlines the slant in the content from these government accounts, making them suitable benchmarks for our measurement.

FIGURE C.1. WORD FREQUENCY IN THE SAMPLE OF GOVERNMENT REPRESENTATIVES' TWEETS



Notes: The figure shows frequencies for selected word stems in the sample of government tweets. In violet, we show the frequency in tweets coming from representatives of the Russian government, and in orange for the Ukrainian government. Frequencies represent the percentage of tweets containing the stem of each specific word of interest. Results are based on 9,451 tweets from Ukrainian government exponents and 5,993 tweets from Russian government exponents.

Following Gennaro and Ash (2023), we take all Ukrainian tweets in the government accounts' tweets, create a vector representation using the text embedding model sentence-t5-xl (Ni et al. 2021), and average those representations to produce a single vector representing the Ukrainian government pole. We compute the Russian government pole analogously. Then, we embed all tweets of our main analysis' dataset (see sample of tweets subsection 4.1) with sentence-t5-xl and use Equation 3 to obtain a score for each input tweet. This score is a ratio measuring the language similarity between the given tweet and the Russian pole relative to the similarity between the given tweet and the Ukrainian pole. Formally, we compute this ratio as follows:

$$Y = \frac{sim(d, R) + b}{sim(d, U) + b} - 1, \tag{3}$$

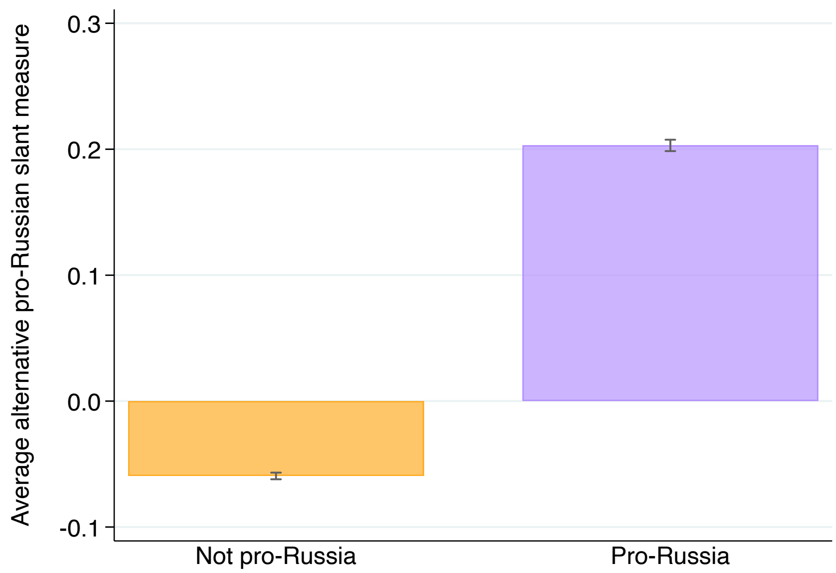
where d denotes a vector representing the input text, R and U are vectors representing the two poles, b is a smoothing parameter set to 1, and sim refers to the cosine similarity. We subtract 1 to center the ratio around zero. Positive values indicate tweets more similar to the Russian pole. Tweets with negative values are more similar to the Ukrainian pole. If $Y = 0$, this means that the input text is equally similar to the Ukrainian and Russian poles. We compute the poles as time-varying measures to account for possible changes over time in the official viewpoints. The comparison poles for a day t consist of all government tweets between days t-7 and t, using a decay factor of 0.5 to reduce the influence of more distant days[21]. The pole ratio for each user tweet from day t is then computed based on the two corresponding daily

[21] The decay factor of 0.5 results in the following weights: 0.5, 0.55204476, 0.60950683, 0.6729501, 0.74299714, 0.82033536, 0.90572366, 1.

poles. Finally, we standardize the resulting media slant ratio to a mean of 0 and a standard deviation of 1. Increasing our final measure by one unit implies moving one standard deviation closer to the Russian pole.

To compare the outcome of this construction with the slant measure we used in the main part, Figure C.2 plots the average continuous pro-Russia slant measure grouped by the binary pro-Russia label used in the main analysis. We document a general agreement between the two measures. On average, we assign higher values of the continuous measure – implying that the tweet leans closer to the Russian government pole – for tweets labeled as pro-Russian slant by our binary measure.

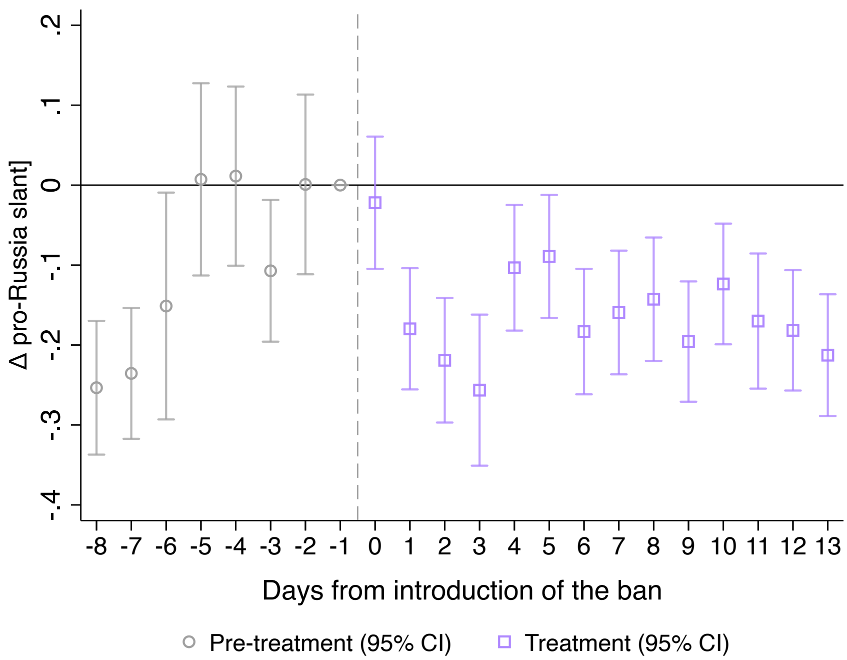
FIGURE C.2. COMPARISON BETWEEN BINARY AND CONTINUOUS PRO-RUSSIA SLANT



Notes: The figure compares the average alternative slant measure for all tweets in our sample grouped by our binary measure of pro-Russia slant.

Finally, in Figure C.3 and Table C.2, we demonstrate that the results presented in the main part of the paper also hold for this alternative, continuous measure of pro-Russian slant. Figure C.3 plots the results of estimating an OLS event-study version of Equation 2 for the continuous slant measure. We find a similar negative shift in pro-Russia content as in Figure II. As in the main figure, the effect reflects the nature of daily social media data and is somewhat noisy. Nonetheless, we find a consistent negative impact of the ban on pro-Russia slant. Column 1 in Table C.2 reproduces the decrease in pro-Russia content presented in Table II. Columns 2 and 3 demonstrate that the finding that the ban's effect was stronger for users connected to the outlets than those not connected, established in the split-sample analysis of Appendix Tables E.1 and E.2, also holds for the alternative slant measure.

FIGURE C.3. DAILY DIFFERENTIAL IMPACT OF THE BAN ON PRO-RUSSIA CONTENT IN EU vs. NON-EU COUNTRIES



Notes: The figure displays coefficients and 95% confidence intervals from estimating an OLS event-study version of Equation 2. The regression uses the user-day panel dataset and includes only user-day observations where the user was active in our sample as described in Section 4.1. We include only observations between February 22^{nd} and March 15^{th}, to ensure a large enough sample size for each day. The dependent variable is the media slant ratio; thus, the coefficients represent the daily average impact of the ban on the level of a user's pro-Russia tweet content, conditional on the user posting at least one tweet on that day. The reported regression includes user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level. The omitted day is March 1^{st}, 2022, the day immediately preceding the introduction of the ban. The vertical dashed line marks the date of the policy intervention.

TABLE C.2.
IMPACT OF THE BAN ON USERS' CONTENT PRODUCTION
ALTERNATIVE MEASURE

Dependent Variable	Total slant		
	All users	**Connected users**	**Not connected users**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Ban × EU	-0.099	-0.162	-0.088
	(0.049)	(0.075)	(0.057)
	[0.045]	[0.031]	[0.123]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓
Pre-ban outcome avg. for treated	-0.258	0.001	-0.327
Observations	322199	57095	265104

Notes: The table presents the results from two-way fixed effects difference-in-difference regressions analyzing the impact of the ban on users' posting behavior by estimating an OLS version of Equation 2. The regression uses the user-day panel dataset and includes only user-day observations where the user was active in our sample as described in Section 4.1. We include only observations between February 22$^{\text{nd}}$ and March 15$^{\text{th}}$, to ensure a large enough sample size for each day. The dependent variable is the media slant ratio; thus, the coefficients represent the daily average impact of the ban on the level of a user's pro-Russia tweet content, conditional on the user posting at least one tweet on that day. Column 1 is analogous to the active-day content-composition specification in Column 4 of Paper Table II, using the alternative continuous measure of pro-Russia slant instead of the discrete measure of pro-Russia tweets. Column 2 shows the effect for users connected to the outlets, reproducing the spirit of the connected split-sample results in Appendix Table E.1. Column 3 presents the effect for non-connected users, reproducing the spirit of the non-connected split-sample results in Appendix Table E.2. All specifications include user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level.

D Triple-Difference Sensitivity: Permutation

This appendix reports a within-country permutation test of the connected variable for the triple-difference estimates. Our baseline clusters standard errors at the user level; Section 5.1 shows robustness to clustering by date, by NUTS2 region, and two-way by user and country-day. All of these approaches, however, rely on asymptotic approximations that are strained when the policy varies across only seven countries. The permutation test provides a complementary, finite-sample-exact inference check that does not rest on any asymptotic argument.

The thought experiment is the following: under the sharp null hypothesis that connection status has no effect on any user's posting behavior, the observed assignment of the connected flag is just one of many equally informative labels, and re-estimating the model under random re-labeling traces out the distribution of the test statistic under that null. Concretely, we hold EU membership and the ban date fixed, and within each country we randomly reassign connected status across users, preserving each country's actual share of connected users. Permuting within country is deliberate: it leaves the country-level structure of the design (EU status, country composition, country-specific shocks) untouched, so the test isolates the user-level dimension on which the triple difference identified: whether *connected* users in EU countries changed differentially after the ban.

For each of 100 permutations we re-estimate our preferred specification – PPML with user and date fixed effects and the pre-ban pro-Russia share and trade exposure controls, each interacted with the post-ban indicator – and record the t-statistic on the permuted triple interaction. We compare absolute t-statistics rather than coefficients, following the randomization-t approach of Young (2019), which is robust to heteroskedasticity in the permutation distribution.

Table D.1 reports the result: no permutation produces an absolute t-statistic as large as the actual value of 3.39, implying a permutation p-value below 0.01. The probability that an arbitrary relabeling of which users are connected would generate an interaction as strong as the one we estimate is thus negligible, confirming that the triple-difference result is not an artifact of the clustering choice or of the small number of countries.

TABLE D.1.
WITHIN-COUNTRY PERMUTATION TEST

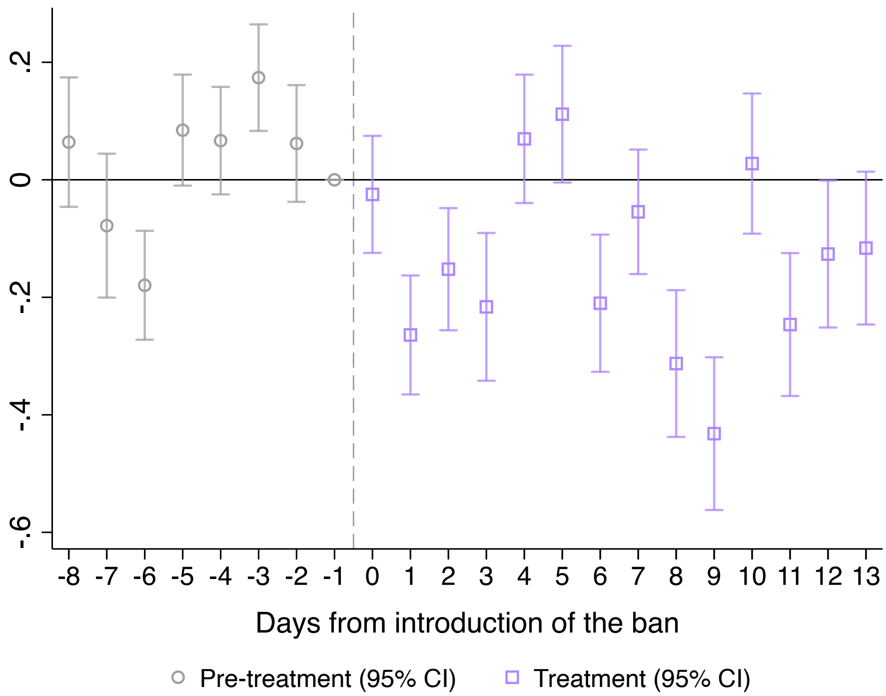
Actual Conn × EU × Ban	-0.2451
Actual t-statistic	-3.39
Permutations	100
Extreme ($\lvert t\rvert \geq \lvert t_{\text{actual}}\rvert$)	0
Permutation p-value	< 0.01

Notes: Within-country permutation test of the connected flag. Holding EU membership and the ban date fixed, connected status is randomly reassigned across users within each country (preserving the country-level share of connected users), and the preferred triple-difference specification of Section 5.1 is re-estimated by PPML for each of 100 permutations. The permutation p-value is the share of permutations with an absolute t-statistic at least as large as the actual $\lvert t\rvert = 3.39$; since no permutation exceeds it, $p < 0.01$.

E Additional Results

In this section, we provide additional results complementing the output we report in the main paper. We start from Figure E.1, which presents the full-sample two-way difference-in-differences event study on the count of pro-Russia tweets per user-day, for EU users relative to non-EU users. This is the full-sample analogue of the triple-difference event study in the main text (Figure II): it pools all users – connected and non-connected – and traces the daily contrast between EU and non-EU around the ban date. The pattern parallels the main result: a sharp downward shift at March 2 in the EU relative to the non-EU users.

FIGURE E.1.
DAILY EVENT STUDY: 2W DiD ON PRO-RUSSIA CONTENT, EU VS. NON-EU

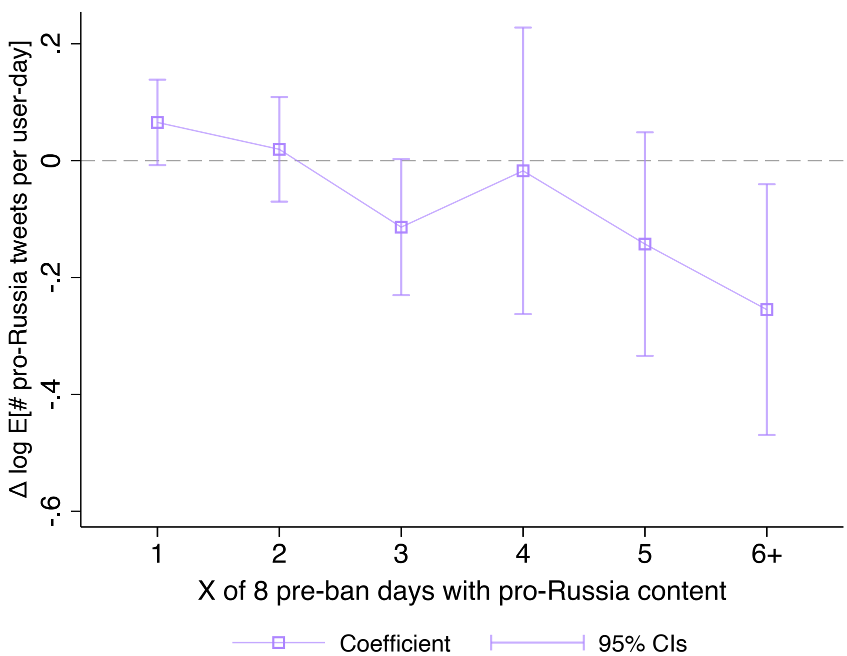


Notes: The figure displays coefficients and 95% confidence intervals from the two-way difference-in-differences event study on the count of pro-Russia tweets per user-day, estimated by PPML on the balanced user-day panel. The model regresses the daily count of pro-Russia tweets on user fixed effects, date fixed effects, and treatment-by-date interactions, with controls for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. We include only observations between February 22nd and March 15th. The omitted day is March 1st, 2022. The vertical dashed line marks the date of the policy intervention. Standard errors are clustered at the user level.

Figure E.2 plots coefficients and 95% confidence intervals from estimating Equation 2 for the impact of the ban on the number of pro-Russia tweets produced by EU-based users relative to non-EU users. We estimate the regressions separately for six subgroups of pro-Russia producers, defined by their level of pre-ban activity: specifically, the total number of days in the eight days before the ban on which a user posted at least one pro-Russia tweet. Subgroup 1 includes users posting on only one day, Subgroup 2 those active on two days, and so on up to Subgroup 6 (the secondary-supplier threshold begins at Subgroup 4). Importantly, the number of users in each group is decreasing in activity, so we top-code the pre-ban days with pro-Russian content at six. We give complete details on the underlying regressions in Appendix Table H.2.

The figure reveals a clear pattern: the ban's negative effect grows in magnitude with the number of days a producer was active in producing pro-Russia content before its implementation, with the largest decline among the most active subgroup (well within the secondary-supplier range) and even a reversal of sign for those active on only one day. While we cannot disentangle the exact mechanism behind this pattern, Appendix Figures B.4 and B.5 provide descriptive context. Panel B.4a shows that the share of producers connected to the outlets' network – either directly or indirectly – is increasing in the number of days with pro-Russian activity before the ban. In Panel B.4b, we additionally show that the average distance of connected producers is decreasing in activity. Therefore, the most active users were also the most embedded in the banned outlets' observed network, consistent with a larger decrease in pro-Russian content. To shed further light on producers and their activity, Figure B.5 plots the average engagement – the sum of retweets, replies, and likes – of pro-Russian content posted by producers separated by their activity. This descriptive output reveals a further dimension: the average pre-ban engagement on pro-Russian content is increasing in activity. Based on the documented overlap between these producer characteristics, we form mutually exclusive combinatorial subgroups for the heterogeneity analysis in the main part of the paper.

FIGURE E.2.
IMPACT ON SECONDARY SUPPLIERS OF PRO-RUSSIA CONTENT
BY LEVEL OF PRE-BAN ACTIVITY



Notes: The figure displays coefficients and 95% confidence intervals from six separate simple difference-in-differences regressions (Equation 2, estimated via PPML) exploring the impact of the ban on pro-Russia producers (users who posted at least one pro-Russia tweet in the eight days before the ban). One regression per subgroup, defined by the user's level of pre-ban activity: from those active on only one day up to those active on at least six days. The secondary-supplier subset corresponds to subgroups 4–6. All regressions use the user-day panel dataset and include only user-day observations where the user posted at least one tweet about the war. We include only observations between February 22nd and March 15th, to ensure a large enough sample size for each day. The dependent variable is the number of pro-Russia tweets posted per user per day; coefficients therefore capture the daily average effect of the ban on pro-Russia activity, conditional on the user being active. As we report raw coefficients, the effects should be interpreted as changes in the logarithm of the expected value of the number of pro-Russia tweets. In Appendix Table H.2, we report complete information on the regressions, including the number of observations and the percentage change for the corresponding difference-in-difference estimation, computed as $(e^{\beta}-1)\cdot 100$. All reported regressions include user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level.

TABLE E.1.
SPLIT-SAMPLE IMPACT OF THE BAN: CONNECTED USERS

Dependent variable	**Any Pro-RU**	**Total Tweets**	**# Pro-RU Tweets**	
Margin	**Extensive**	**Volume**	**Unrestricted**	**Cond. on Active Day**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban $\times$ EU	-0.168	-0.139	-0.232	-0.178
	(0.037)	(0.039)	(0.061)	(0.059)
	[0.000]	[0.000]	[0.000]	[0.003]
Unconditional	✓	✓	✓	
Conditional on Posting				✓
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.263	1.259	0.498	1.083
Approx. Percentage Change	-15.44	-12.94	-20.68	-16.28
Observations	145024	145024	145024	50458

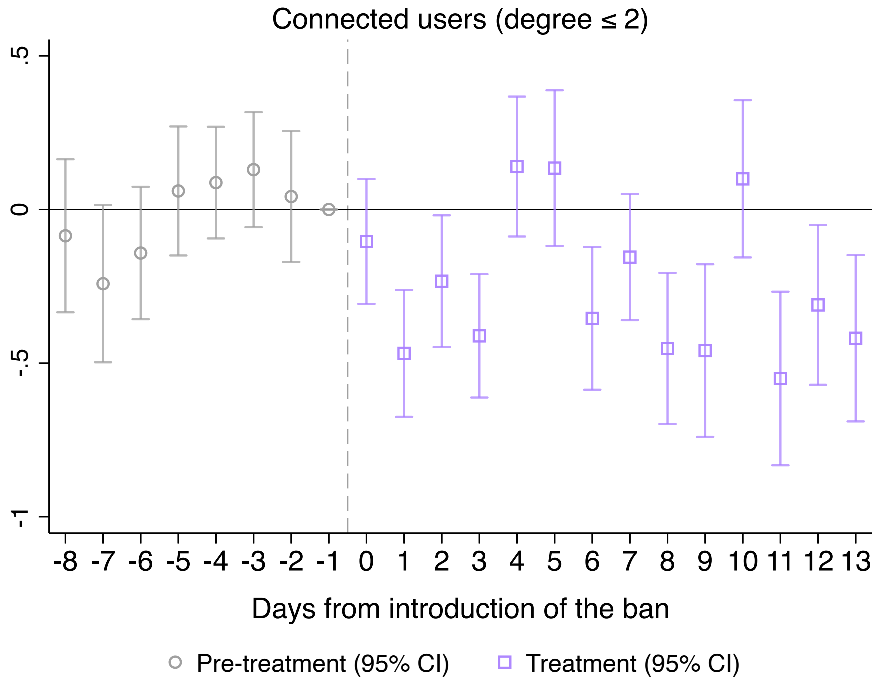
Notes: Four-outcome PPML specification (mirroring main Table II), estimated on the connected subsample (users within network degree ≤ 2 of RT or Sputnik, or who interacted directly with either outlet before the ban). User and date fixed effects, and controls for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. Standard errors clustered at the user level.

TABLE E.2.
SPLIT-SAMPLE IMPACT OF THE BAN: NON-CONNECTED USERS

Dependent variable	**Any Pro-RU**	**Total Tweets**	**# Pro-RU Tweets**	
Margin	**Extensive**	**Volume**	**Unrestricted**	**Cond. on Active Day**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	-0.012	-0.043	-0.100	-0.077
	(0.022)	(0.043)	(0.035)	(0.027)
	[0.586]	[0.310]	[0.004]	[0.004]
Unconditional	✓	✓	✓	
Conditional on Posting				✓
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.116	0.462	0.158	0.641
Approx. Percentage Change	-1.21	-4.25	-9.49	-7.39
Observations	996490	996490	996490	165265

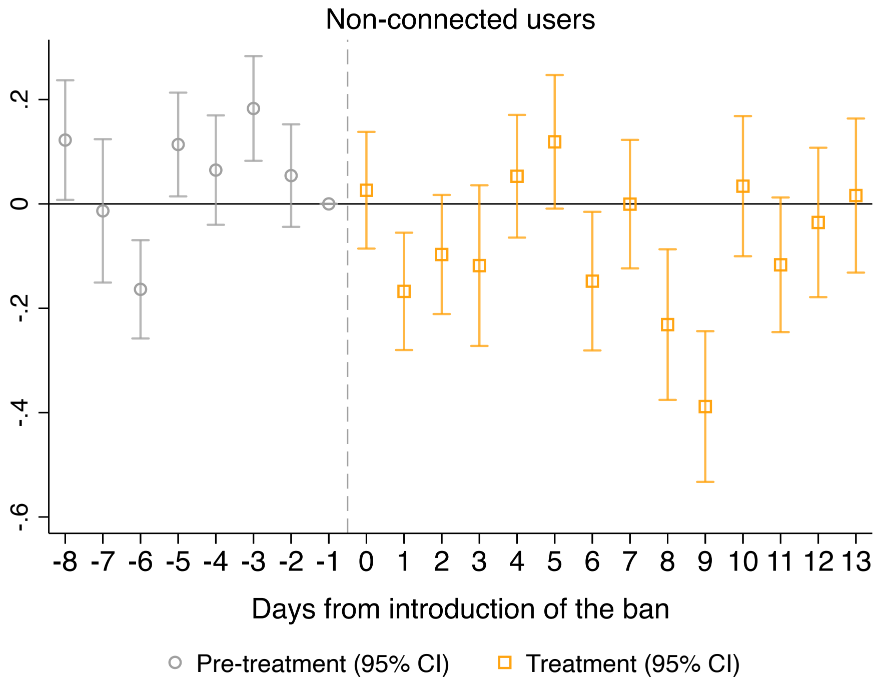
Notes: Same specification as Table E.1, estimated on the non-connected subsample (users with no pre-ban interaction with RT or Sputnik and beyond network degree 2). Effects are smaller in magnitude than the connected panel, consistent with the supply-disruption reading.

FIGURE E.3.
SPLIT-SAMPLE EVENT STUDY: CONNECTED USERS



Notes: Daily event study of the EU-by-day interaction on the connected subsample, estimated via PPML with user and date fixed effects and controls for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. The omitted reference day is March 1, 2022 – the last day before the ban took effect.

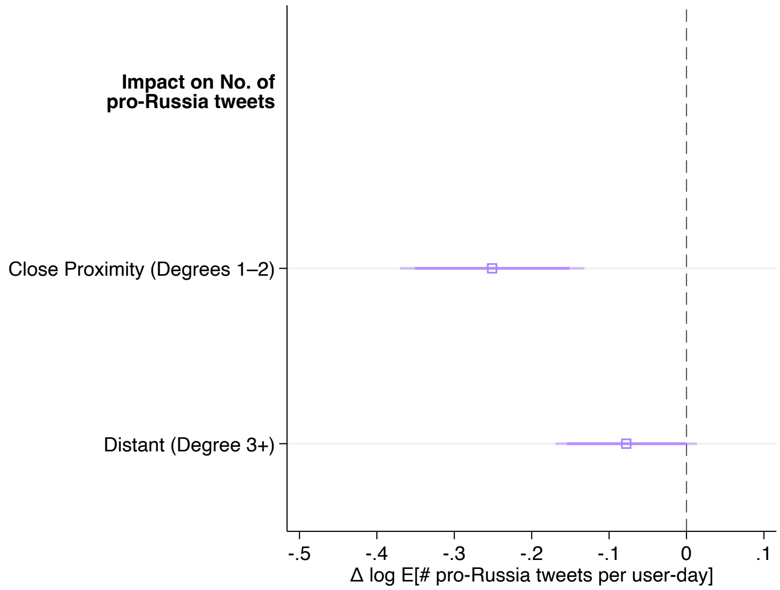
FIGURE E.4.
SPLIT-SAMPLE EVENT STUDY: NON-CONNECTED USERS



Notes: Daily event study on the non-connected subsample. Same specification as Figure E.3.

Figure E.5 examines whether the ban's effect varies with a user's pre-ban network proximity to the banned outlets. We measure proximity as the shortest-path distance to the combined RT/Sputnik node in the retweet-and-reply network constructed from all pre-ban tweets (February 16 – March 2, 2022). A distance of one corresponds to users who directly retweeted or replied to RT/Sputnik, a distance of two to users who interacted with a direct interactor, and so on. For each group separately, we estimate our baseline Poisson difference-in-differences specification. The difference between the two coefficients is the main treatment effect estimated in Column 3 of Table I. Figure E.5 and the accompanying Table H.1 unpack the triple-difference result into separate difference-in-difference effects for each group, thereby allowing for differences in the fixed effects between the connected and unconnected groups of users. On top of the differential effect of the ban on the connected and unconnected users within the EU, established by the triple-difference estimation, the two separate difference-in-difference effects highlight the level of the effect within each group. The decline in the distant user group is consistent with the ban affecting the indirect diffusion of RT/Sputnik content.

FIGURE E.5.
HETEROGENEOUS IMPACT OF THE BAN BY PROXIMITY TO THE OUTLETS



Notes: Coefficient plot of the by-proximity heterogeneity, PPML estimates on the full balanced user-day panel (matching Column 3 of Table II). The figure visualizes the same regressions whose point estimates and stats are reported in Table H.1.

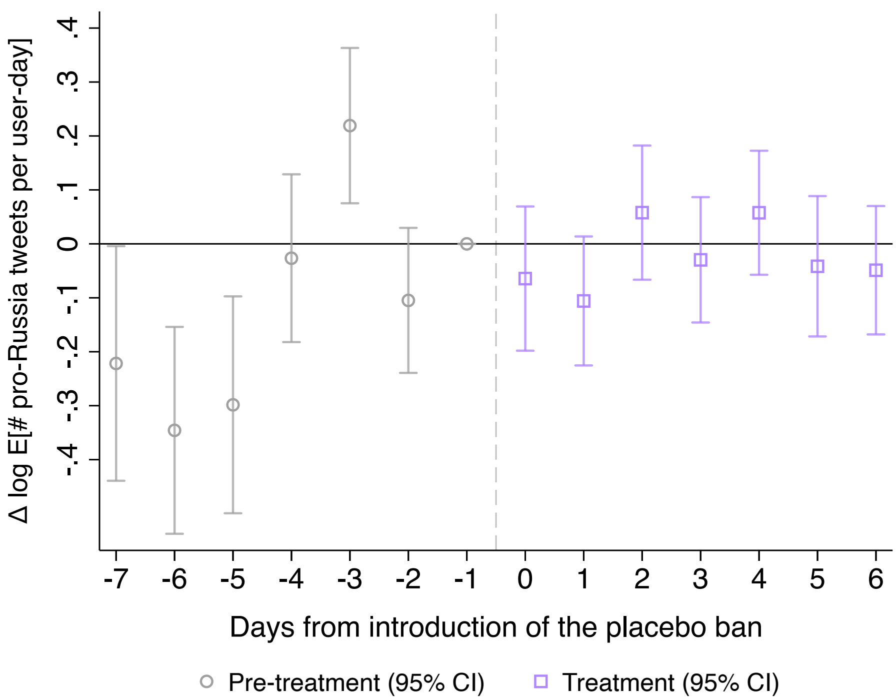
F Placebo Checks

In this section, we present three additional checks that probe the internal validity of our natural experiment: a placebo test on the timing of the intervention, a placebo outcome (anti-Russia content), and a newspaper-based test for confounding shifts in European media coverage. We begin with the timing placebo. To our knowledge, the setting analyzed here – a top-down, large-scale media blackout – is unique in the literature, so it is essential to verify that the effects we document are driven by the ban itself rather than by pre-existing, idiosyncratic differences between the EU and non-EU samples. Figure F.1 replicates the specification used in Figure II, but shifts the break point to a fictitious ban date five days before the actual implementation. If our identification strategy is valid, the coefficients plotted after this placebo date should display no systematic change.

The results are reassuring and convey a clear message: the differences identified in our main analysis are driven by the ban itself. In the days that follow the placebo date, the coefficients show no differential effect between EU and non-EU users. One caveat is that this placebo specification extends the pre-treatment window, adding several dates we normally exclude because of the limited number of observations and their distance from both the invasion and the actual ban; this helps explain the noisier pre-treatment coefficients in the extended placebo window. Overall, the evidence supports the validity of our identification strategy.

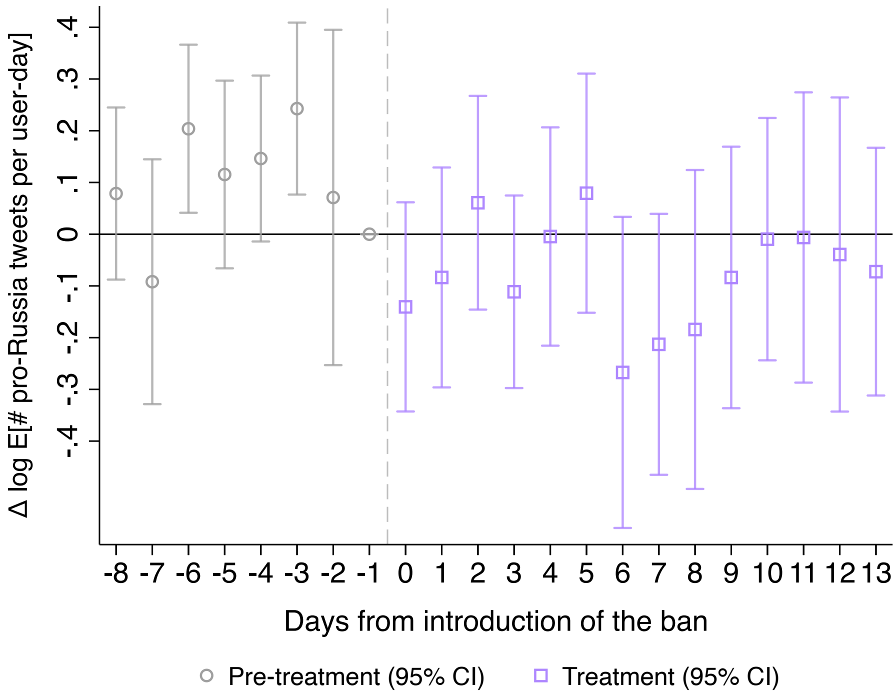
The second placebo exercise, reported in Figure F.2, investigates whether the ban influenced anti-Russia content. As a reminder, our classification pipeline assigns every tweet a value label: pro-Russia when the text aligns with Kremlin narratives, neutral, or anti-Russia when it opposes them. The core analysis focuses on pro-Russia tweets to test whether the ban cut their prevalence in the EU. Within our conceptual framework, there is no reason to expect a symmetric effect on anti-Russia messages, unless those posts were produced solely in response to pro-Russia content. We instead view users as actors who seek to spread their own worldview independently of their counterparts. Consistent with that view, Figure F.2 shows that the ban leaves the volume of anti-Russia tweets essentially unchanged.

FIGURE F.1.
DAILY EVENT STUDY: IMPACT OF THE PLACEBO BAN ON PRO-RUSSIA CONTENT IN EU VS. NON-EU



Notes: The figure displays coefficients and 95% confidence intervals from estimating the event-study version of Equation 2. We use a placebo event, shifting the implementation five days earlier than the actual ban, i.e. from March 2$^{\text{nd}}$ to February 25$^{\text{th}}$, 2022. The regression uses the balanced user-day panel and includes observations between February 16$^{\text{th}}$ and March 1$^{\text{st}}$, to ensure a large enough sample size for each day. The dependent variable is the number of pro-Russia tweets posted per user per day; thus, the coefficients capture the daily average effect of the placebo event on expected pro-Russia tweet activity per user-day. As we report raw coefficients, the effects should be interpreted as changes in the logarithm of the expected value of the number of pro-Russia tweets. The reported regression includes user and day fixed effects, and standard errors are clustered at the user level. The omitted day is February 24$^{\text{th}}$, 2022, the day immediately preceding the introduction of the placebo ban. The vertical dashed line marks the date of the placebo policy intervention.

FIGURE F.2.
DAILY EVENT STUDY: IMPACT OF THE BAN ON ANTI-RUSSIA CONTENT IN EU VS. NON-EU



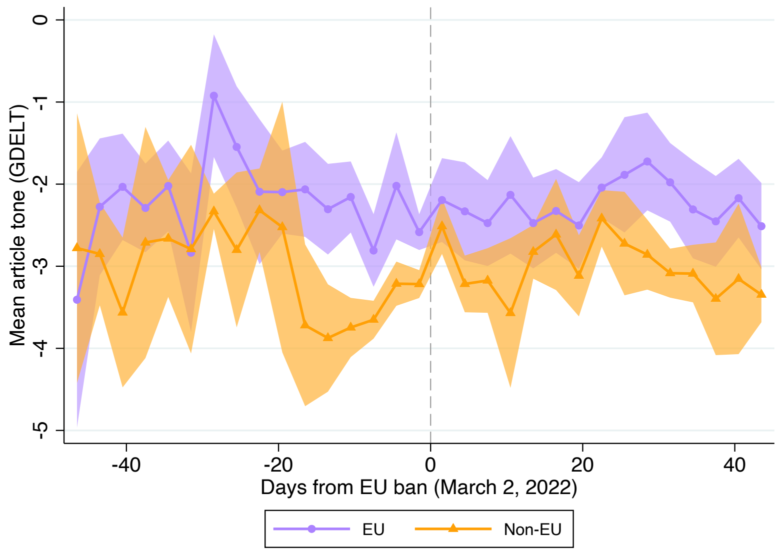
Notes: The figure displays coefficients and 95% confidence intervals from estimating the event-study version of Equation 2. The regression uses the balanced user-day panel and includes observations between February 22$^{\text{nd}}$ and March 15$^{\text{th}}$, to ensure a large enough sample size for each day. The dependent variable is the number of anti-Russia tweets posted per user per day; thus, the coefficients capture the daily average effect of the ban on expected anti-Russia tweet activity per user-day. As we report raw coefficients, the effects should be interpreted as changes in the logarithm of the expected value of the number of anti-Russia tweets. The reported regression includes user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level. The omitted day is March 1$^{\text{st}}$, 2022, the day immediately preceding the introduction of the ban. The vertical dashed line marks the date of the policy intervention.

Newspaper coverage around the ban. A remaining concern is that the post-ban decline in pro-Russia content among EU users reflects a broader shift in how European media covered the war, rather than the platform-specific removal of RT and Sputnik. Newspapers provide a natural placebo outlet: the ban applied to the broadcasting and online dissemination of the two sanctioned outlets, not to domestic press coverage of Russia. If EU public discourse as a whole turned against Russia at the ban date for reasons unrelated to the ban, this shift should be visible in newspapers on both sides of the EU border; if instead the Twitter result reflects the censorship intervention, newspaper coverage should evolve in parallel across EU and non-EU countries.

We test this using GDELT, which monitors online news worldwide and assigns each article a tone score from a dictionary-based sentiment analysis of the full text (negative values indicate more negative coverage). We build an outlet–day panel of Russia–Ukraine coverage for 16 major daily outlets in the five largest EU countries of our Twitter sample (Germany, France, Italy, Austria, Ireland) and the two non-EU control countries (United Kingdom, Switzerland), identifying relevant articles through language-specific keyword filters (e.g., Russia, Ukraine, Putin, Kremlin, NATO, sanctions, and their German, French, and Italian equivalents). For the 2022 window (January 15–April 15, 2022) we query the GDELT DOC API for each outlet's daily average tone; for the 2014 placebo window (January 1–June 30, 2014) we process the underlying GDELT Global Knowledge Graph files directly, as the API's article archive does not extend that far back. Tone is undefined on days when an outlet publishes no matching article, so these outlet-days are excluded. For the figures, we first average within country and then across countries, so that countries with many sampled outlets do not dominate the EU aggregate, and we smooth daily noise using 3-day bins.

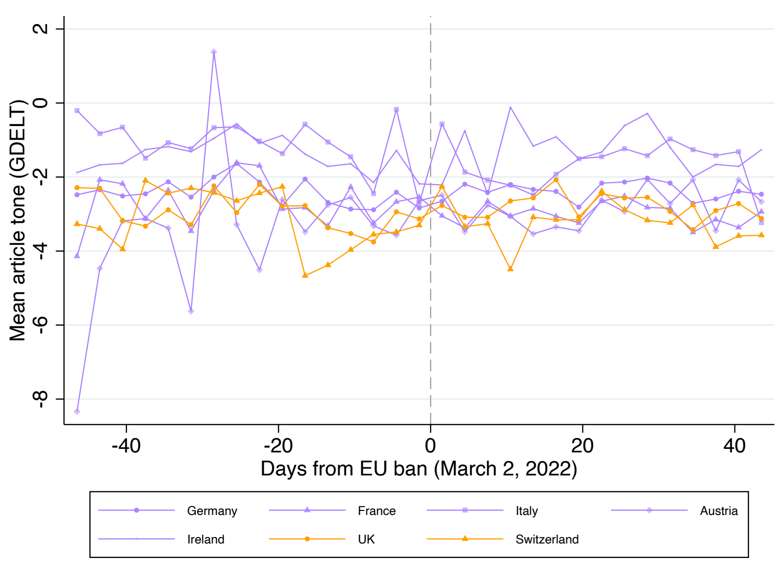
Figure F.3 shows that EU and non-EU newspaper tone moves in lockstep throughout the window: both groups turn sharply more negative at the February 24 invasion, and neither shows any visible break at the March 2 ban. The country-level disaggregation in Figure F.4 confirms that this pattern is not an artifact of aggregation. Table F.1 quantifies this in a difference-in-differences design with outlet fixed effects and standard errors clustered at the country level: the EU $\times$ Post coefficient on tone is 0.02 (s.e. 0.15)—less than two percent of the outcome's residual standard deviation—and insensitive to day-of-week controls. Finally, the 2014 Crimea annexation, a comparable Russian geopolitical shock that triggered no media ban, serves as a placebo event: Figure F.5 and Table F.2 show no differential EU-versus-non-EU tone response around March 18, 2014 either. We note that GDELT's archival coverage of continental European outlets in 2014 is sparse, so the placebo comparison rests primarily on English-language outlets (Irish versus British newspapers) and only three country clusters; we therefore read the 2014 evidence from the event-study figure rather than leaning on the regression's small-cluster inference. Taken together, the newspaper evidence indicates that the decline in pro-Russia content we document on Twitter is specific to the platform environment targeted by the ban, and does not reflect a generalized shift in European media coverage of the war.

FIGURE F.3.
GDELT NEWSPAPER PLACEBO: EU vs. NON-EU TONE AROUND THE BAN (2022)



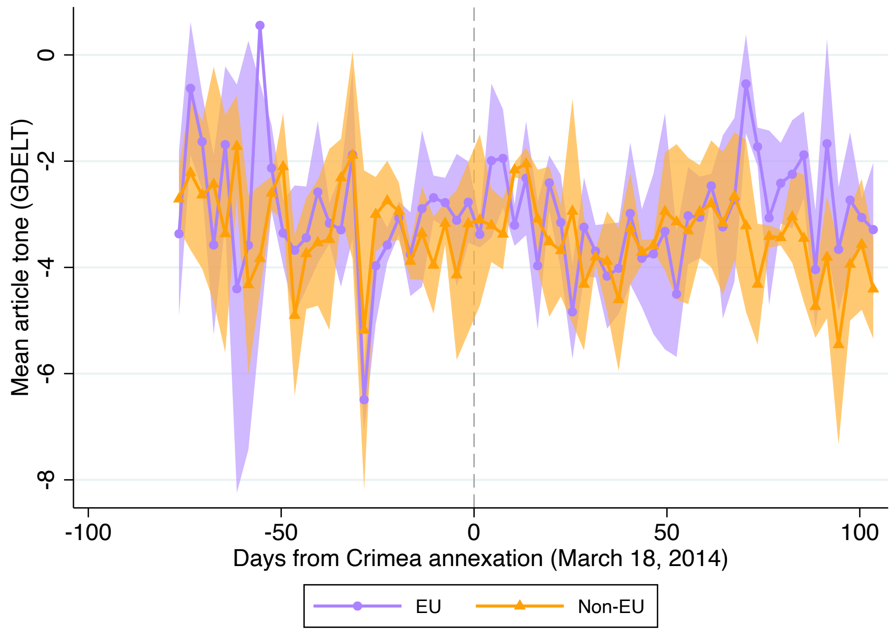
Notes: Mean article tone (GDELT) for EU vs. non-EU outlets, binned in 3-day windows around the March 2, 2022 ban date (dashed line). Shaded bands are ±1.96 standard errors of the bin mean computed across country–day observations within each bin. Both groups turn sharply more negative at the February 24 invasion; no differential shift is visible at the ban. Complements Table F.1.

FIGURE F.4.
GDELT NEWSPAPER PLACEBO: TONE BY COUNTRY (2022)



Notes: Disaggregation of Figure F.3 by country (3-day bins). No country exhibits a systematic shift at the March 2, 2022 ban date relative to its pre-ban trajectory.

FIGURE F.5.
GDELT CRIMEA PLACEBO: EU vs. NON-EU TONE AROUND MARCH 18, 2014



Notes: Placebo using the 2014 Crimea annexation (March 18, 2014, dashed line) as the event date; 3-day bins. Shaded bands are ±1.96 standard errors of the bin mean computed across country–day observations within each bin. GDELT coverage of continental European outlets in 2014 is sparse, so the comparison is predominantly between Irish (EU) and British (non-EU) outlets. No differential shift is visible, indicating that the 2022 analysis does not capture a mechanical European-media pattern around Russian geopolitical shocks.

TABLE F.1.
GDELT NEWSPAPER DiD (2022): EU × Post

	(1) Tone	(2) Tone + DoW
EU × Post	0.0195	0.0186
	(0.153)	(0.154)
N	1024	1024

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Outlet–day DiD on GDELT article tone around the March 2, 2022 ban. Outlet fixed effects; standard errors clustered at the country level (7 clusters); inference with few clusters should be interpreted with caution. Tone is defined only on outlet-days with at least one matching article (N = 1,024). Column (2) adds day-of-week fixed effects.

TABLE F.2.
GDELT NEWSPAPER DiD (2014 CRIMEA PLACEBO)

	(1) Tone	(2) Tone + DoW
EU $\times$ Post	0.257	0.257
	(0.233)	(0.234)
N	1069	1069

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Same specification as Table F.1, with the 2014 Crimea annexation (March 18, 2014) as the placebo event. GDELT coverage of continental European outlets in 2014 is sparse: the estimation sample comprises 7 outlets in 3 country clusters, predominantly English-language (Irish vs. British) newspapers. With 3 clusters the clustered variance matrix is near-singular and the clustered standard error in column (1) is unreliable; the table is reported for completeness and the placebo conclusion rests on Figure F.5.

G Alternative Pro-Russia Source: TASS

We explore a complementary question to our placebo exercises: Is there evidence of a substitution effect among alternative propaganda channels from Russia? In other words, did the ban merely silence specific actors, or did it also reshape the observed media environment by prompting users to seek out, or step into, alternative sources of pro-Russia content? The answer matters. If engagement with such content remained stable in the short term, the removal of key suppliers may have opened up a vacuum. Below, we explore whether such substitution dynamics emerged in the two weeks following the ban.

There is not a unique type of user that could have substituted the outlets as sources of propaganda; nevertheless, here, we focus on an "institutional candidate", a user as close as possible to the banned outlets, namely TASS, the Russian News Agency. We examine the effect of the ban on TASS, as reported in Table G.1. For this analysis, we rely on the dataset covering the Twitter activity of the outlets. As a reminder, we have complete tweet-level data for Russia Today and Sputnik, and we collected the same information for TASS. In the table, we implement a descriptive contrast following Equation 2, with TASS compared to the targeted benchmark outlets, Russia Today and Sputnik. The coefficients should be interpreted as the differential post-ban change for TASS relative to RT/Sputnik, not as a causal comparison to a separate untreated benchmark. To avoid confusion, it is worth clarifying that Russia Today and Sputnik remained able to post content during the period of analysis; however, their tweets were blocked from being accessed by users within the EU. Users located outside the EU could still view and interact with their content as usual.

The table presents two types of outcome variables. Columns 1 to 3 use the number of pro-Russia tweets as the dependent variable. It is important to note that these outlets often present themselves as neutral or reliable news sources. However, when they do post pro-Russia content, the messaging tends to be strongly aligned with the Russian government's worldview, often taking a more extreme tone. The second type of dependent variable, shown in Columns 4 to 6, is the engagement received by pro-Russia tweets, measured as the sum of retweets, likes, and replies. It should be interpreted as a measure of engagement, conditional on the content being classified as pro-Russia. Columns 1 and 4 capture the most basic structure, with no fixed effects. Columns 2 and 5 add a trend capturing the level of activity; columns 3 and 6 further add day fixed-effects.

The results paint a clear picture: the ban did not statistically significantly affect TASS' activity or engagement levels relative to Russia Today and Sputnik. The point estimates suggest, if anything, that TASS reduced its pro-Russia output post-ban, but none of the estimated models yield statistically significant differences. For engagement, the point estimates are positive in Columns 4 to 6, rise slightly when adding the activity trend, and then fall once day fixed effects are included; none is statistically distinguishable from zero. Overall, we find no evidence of institutional substitution within this descriptive contrast. The void left by Russia Today and Sputnik does not appear to have been filled by another state-backed source like TASS in the two-week window we observe.

TABLE G.1.
IMPACT OF THE BAN ON THE RUSSIAN NEWS AGENCY: TASS

Dependent variable	**# tweets pro-Russia**			**Engagement**		
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
TASS vs. Russia Today and Sputnik	-0.339 (0.222) [0.128]	-0.328 (0.222) [0.140]	-0.313 (0.221) [0.157]	0.278 (0.203) [0.173]	0.304 (0.203) [0.133]	0.191 (0.235) [0.417]
Daily activity trend		✓	✓		✓	✓
Day FE			✓			✓
Pre-ban avg. outcome for TASS	0.298	0.298	0.298	87.765	87.765	87.765
Approx. percentage change	-28.72	-27.97	-26.86	31.99	35.57	21.07
Observations	4651	4651	4651	853	853	853

Notes: The table examines TASS' activity and the engagement it received after the ban, using RT/Sputnik as the targeted benchmark outlets. It reports results from Poisson-Pseudo Maximum Likelihood regressions estimated in a difference-in-difference framework at the tweet level. The coefficients capture the differential post-ban change for TASS relative to Russia Today and Sputnik, which remained active after the policy was introduced but were the direct targets of the EU geo-block. They should therefore be read as descriptive contrasts, not as estimates against a separate untreated benchmark. Columns 1 to 3 estimate the effect on the volume of pro-Russia content produced, while Columns 4 to 6 estimate the effect on total retweets, replies, and likes received by pro-Russia tweets. All models use data from February 22$^{\text{nd}}$ to March 15$^{\text{th}}$, 2022. Percentage changes are computed as $e^{\beta} - 1$.

H Supporting Regression Tables

This section provides the underlying regression models for the heterogeneity coefficient plots reported later in this appendix. Table H.1 reports the by-proximity regressions visualized in Figure E.5; Table H.2 reports the by-activity regressions visualized in Appendix Figure E.2; Table H.3 reports the activity × engagement regressions visualized in Figure IV, with the corresponding engagement-outcome regressions in Table H.4 and Figure V.

TABLE H.1.
HETEROGENEOUS IMPACT OF THE BAN BY PROXIMITY TO THE OUTLETS: UNDERLYING MODELS

Dependent Variable	**Total No. of Pro-Russia Tweets**	
Proximity	**Degree 1-2**	**Degree 3+**
	Coeff./SE/p-value	
	(1)	(2)
Ban × EU	-0.251 (0.061) [0.000]	-0.078 (0.047) [0.096]
User FEs	✓	✓
Date FEs	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓
Pre-Ban Trade with Russia: State	✓	✓
Pre-Ban Outcome Avg. for Treated	0.477	0.215
Approx. Percentage Change	-22.20	-7.49
Observations	153010	461934

Notes: The table displays the underlying simple difference-in-differences regressions behind Figure E.5, estimated via Poisson Pseudo-Maximum Likelihood (PPML; i.e. Equation 2). We present results from two separate regressions that share an identical specification but differ in the user samples analyzed. Both regressions use the full balanced user-day panel (matching Column 3 of Table II); we include only observations between February 22^{nd} and March 15^{th}, to ensure a large enough sample size for each day. The coefficient in column 1 shows the impact of the ban on users with high proximity to Russia Today and Sputnik, defined as users who directly retweeted or replied (first-degree connection) to posts from these outlets or interacted with posts by users who had a first-degree connection. In column 2, the coefficient reflects the impact on users with weaker connections, defined as third-degree proximity or higher. The approximate percentage change is computed as $(e^{\beta} - 1) \cdot 100$. All regressions include user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level.

TABLE H.2.
HETEROGENEOUS IMPACT OF THE BAN ON PRO-RUSSIA CONTENT FOR EACH LEVEL OF PRE-BAN ACTIVITY: UNDERLYING MODELS

Dependent Variable	**Total No. of Pro-Russia Tweets**					
Days posting Pro-Russia	**1**	**2**	**3**	**4**	**5**	**6+**
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Ban × EU	0.065	0.019	-0.114	-0.017	-0.143	-0.255
	(0.037)	(0.046)	(0.059)	(0.125)	(0.098)	(0.109)
	[0.079]	[0.672]	[0.056]	[0.889]	[0.143]	[0.020]
User FEs	✓	✓	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓	✓	✓
Conditional on Posting About War	✓	✓	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓	✓	✓
Pre-Ban Outcome Avg. for Treated	0.613	0.937	1.242	1.805	1.969	3.696
Approx. Percentage Change	6.76	1.96	-10.76	-1.73	-13.30	-22.50
Observations	60024	26949	13705	7483	4581	4989

Notes: The table displays the underlying simple difference-in-differences regressions behind Appendix Figure E.2, estimated via Poisson Pseudo-Maximum Likelihood (PPML; i.e. Equation 2), exploring the impact of the ban on pro-Russia producers. Users are defined as pro-Russia producers if they posted at least one pro-Russia tweet in the eight days before the ban; the secondary-supplier subset corresponds to columns 4–6 (active on ≥ 4 of 8 pre-ban days). We run six separate regressions, one for each subgroup defined by the user's level of pre-ban activity: from those active on only one day up to those active on at least six days. All regressions use the user-day panel dataset and include only user-day observations where the user posted at least one tweet about the war. We include only observations between February 22$^{\text{nd}}$ and March 15$^{\text{th}}$, to ensure a large enough sample size for each day. The dependent variable is the number of pro-Russia tweets posted per user per day; coefficients therefore capture the daily average effect of the ban on pro-Russia activity, conditional on the user being active. As we report raw coefficients, the effects should be interpreted as changes in the logarithm of the expected value of the number of pro-Russia tweets. The approximate percentage change is computed as $(e^{\beta} - 1) \cdot 100$. All reported regressions include user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level.

TABLE H.3.
IDENTIFYING THE USERS MOST AFFECTED BY THE BAN: UNDERLYING MODELS

Dependent Variable	**Total No. of Pro-Russia Tweets**			
Samples	**Low Act., Regular Engag.**	**Low Act., Top Engag.**	**High Act., Regular Engag.**	**High Act., Top Engag.**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	0.017	0.100	-0.062	-0.269
	(0.043)	(0.131)	(0.091)	(0.127)
	[0.690]	[0.446]	[0.491]	[0.033]
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Outcome Avg. for Treated	0.253	0.498	1.324	2.819
Approx. Percentage Change	1.73	10.47	-6.05	-23.62
Observations	488334	19844	21692	8162

Notes: The table displays the underlying simple difference-in-differences regressions behind Figure IV, estimated via Poisson Pseudo-Maximum Likelihood (PPML; i.e. Equation 2). All regressions use the full balanced user-day panel (matching Column 3 of Table II); we include only observations between February 22$^{\text{nd}}$ and March 15$^{\text{th}}$, to ensure a large enough sample size for each day. The figure divides pre-ban pro-Russia producers by activity – low (1–3 days posting pro-Russia content) vs. high (4–8 days, the secondary-supplier subset) – and by the engagement their pro-Russia content received before the ban (regular vs. top 5% by retweets, likes, and replies). The dependent variable is the number of pro-Russia tweets posted per user per day. As we report raw coefficients, the effects should be interpreted as changes in the logarithm of the expected value of the number of pro-Russia tweets. The approximate percentage change is computed as $(e^{\beta} - 1) \cdot 100$. All reported regressions include user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level.

TABLE H.4.
ENGAGEMENT OF SECONDARY SUPPLIERS' PRO-RUSSIA CONTENT: UNDERLYING MODELS

Dependent Variable	**Engagement on Pro-Russia Tweets (per user-day)**			
Samples	**Low Act., Regular Engag.**	**Low Act., Top Engag.**	**High Act., Regular Engag.**	**High Act., Top Engag.**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	0.377 (0.225) [0.093]	0.042 (0.312) [0.894]	-0.215 (0.212) [0.311]	0.199 (0.206) [0.334]
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Conditional on Posting About War	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Outcome Avg. for Treated	2.040	110.567	3.044	108.075
Approx. Percentage Change	45.75	4.26	-19.33	22.01
Observations	77577	6528	11432	5477

Notes: The table displays the underlying simple difference-in-differences regressions behind Figure V, estimated via Poisson Pseudo-Maximum Likelihood (PPML; i.e. Equation 2). The dependent variable is per-user-day engagement on pro-Russia content (sum of likes, retweets, and replies on a user's pro-Russia tweets that day); the PPML coefficient interprets as the log-multiplier on expected per-user-day engagement. The activity × engagement partition, fixed effects, and controls (pre-ban pro-Russia attitudes and local trade exposure, each interacted with the ban) match Table H.3; the sample is restricted to user-days with at least one tweet about the war (active-day sample), since per-tweet engagement is undefined when no tweets are posted. The approximate percentage change is computed as $(e^{\beta} - 1) \cdot 100$. Standard errors are clustered at the user level.

I Robustness Checks

This appendix reports three robustness exercises for the full-sample two-way DiD results: excluding potential bots, excluding accounts created after the ban, and re-estimating the main specifications with OLS rather than PPML.

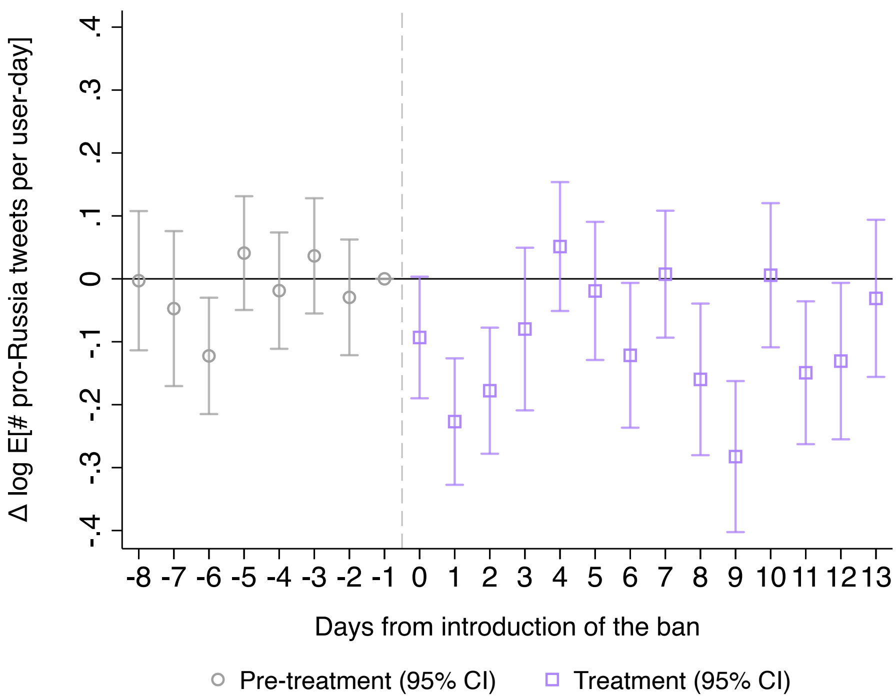
I.1 Potential Bots

In this subsection, we replicate the main findings of our analysis, adjusting our sample by removing potential bots. There is not one ideal or optimal way to identify potential bots. To identify such profiles, we rely on the criteria established by studies from the computer science literature, that suggest that Twitter bots typically exhibit a high frequency of tweets per day (Tabassum et al. 2023) and a low 'reputation' ratio, calculated as the number of followers divided by the sum of the number of followers and the number of accounts followed (Chu et al. 2012). Following the recommendations of Gehring and Grigoletto (2023), we classify potential bots as accounts ranking in the upper 25% for daily tweet frequency and in the lower 25% for the reputation metric. In addition, we exclude high-volume pre-ban pro-Russia users – those who posted more than 80 pro-Russia tweets in the eight days preceding the ban (an average above ten per day). This threshold is intended to capture plausible bot-like activity, but it also removes genuine high-volume activists whose response to the ban could be substantively interesting.

Excluding these users therefore biases the estimates against finding a strong supplier-side response, which makes any remaining effect conservative for that subgroup. Applying these criteria leads to the exclusion of 2,585 users.

In Figure I.1 and Table I.1, we reproduce the main results of our analysis. The exclusion of these users does not substantially alter the effects of the ban, suggesting that bots are not a driver in the analysis. Naturally, our approach may not allow us to pinpoint the right profiles. Nevertheless, if undetected Russian bots attempted to counteract the effect of the ban, the effects we find may be attenuated.

FIGURE I.1.
DAILY EVENT STUDY: IMPACT OF THE BAN ON PRO-RUSSIA CONTENT IN EU VS. NON-EU EXCLUDING POTENTIAL BOTS



Notes: The figure displays coefficients and 95% confidence intervals from estimating the event-study version of Equation 2. It reproduces the two-way event-study specification in Appendix Figure E.1, excluding potential bots. The regression uses the balanced user-day panel for users in the sample. A potential bot is identified by either one of the following conditions: The user is in the bottom 25% of reputation and top 25% of tweet production, or the user produced more than 80 pro-Russia tweets in the eight days before the ban. We include only observations between February 22nd and March 15th, to ensure a large enough sample size for each day. The dependent variable is the number of pro-Russia tweets posted per user per day; thus, the coefficients capture the daily average effect of the ban on expected pro-Russia tweet activity per user-day. As we report raw coefficients, the effects should be interpreted as changes in the logarithm of the expected value of the number of pro-Russia tweets. The reported regression includes user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level. The omitted day is March 1st, 2022, the day immediately preceding the introduction of the ban. The vertical dashed line marks the date of the policy intervention.

TABLE I.1.
IMPACT OF THE BAN ON PRO-RUSSIA AND OVERALL WAR-RELATED CONTENT EXCLUDING POTENTIAL BOTS

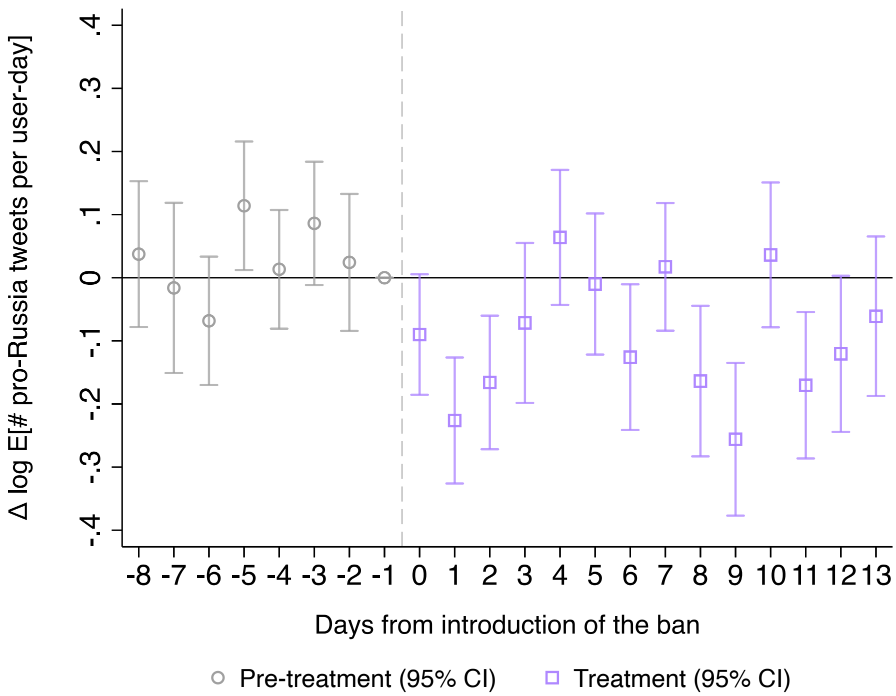
Dependent variable	**Any Pro-RU**	**Total Tweets**	**# Pro-RU Tweets**	
Margin	**Extensive**	**Volume**	**Unrestricted**	**Cond. on Active Day**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	-0.052 (0.020) [0.008]	-0.056 (0.037) [0.130]	-0.126 (0.030) [0.000]	-0.078 (0.023) [0.001]
Unconditional	✓	✓	✓	
Conditional on Posting				✓
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.137	0.561	0.200	0.733
Approx. Percentage Change	-5.04	-5.44	-11.87	-7.48
Observations	1118018	1118018	1118018	209494

Notes: The table re-estimates the main Table II specification on the sample excluding potential bots. Bots are users in the bottom 25% of the follower-to-followee reputation metric and the top 25% of daily tweet production, plus users who posted more than 80 pro-Russia tweets in the eight days before the ban. All four columns are estimated via Poisson Pseudo-Maximum Likelihood (PPML). The coefficient of interest is the Ban × EU interaction. All specifications include user and date fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. Outcomes across columns match Table II: Column 1 is an indicator for posting any pro-Russia content on a given day (extensive margin); Column 2 is total tweet volume; Column 3 is the count of pro-Russia tweets on the balanced panel; Column 4 restricts to days with at least one tweet and captures compositional shift on active days. Percentage changes are computed as $(e^{\beta} - 1) \cdot 100$. Pre-ban means and percentage changes are computed on each column's estimation sample. Standard errors are clustered at the user level.

I.2 Accounts Created After the Ban

In this subsection, we replicate the main findings of our analysis, adjusting our sample by removing accounts that were created after the ban. We do this to shed light on whether the creation of new accounts was an important mechanism of reaction to the ban. The exercise leads to the removal of 458 profiles. Like for potential bots, also in this case the removal of these profiles does not substantially affect results, as shown in Figure I.2 and Table I.2.

FIGURE I.2.
DAILY EVENT STUDY: IMPACT OF THE BAN ON PRO-RUSSIA CONTENT IN EU VS. NON-EU EXCLUDING ACCOUNTS CREATED POST-BAN



Notes: The figure displays coefficients and 95% confidence intervals from estimating the event-study version of Equation 2. It reproduces the two-way event-study specification in Appendix Figure E.1, excluding accounts that were created after the ban. The regression uses the balanced user-day panel for users in the sample. We include only observations between February 22nd and March 15th, to ensure a large enough sample size for each day. The dependent variable is the number of pro-Russia tweets posted per user per day; thus, the coefficients capture the daily average effect of the ban on expected pro-Russia tweet activity per user-day. As we report raw coefficients, the effects should be interpreted as changes in the logarithm of the expected value of the number of pro-Russia tweets. The reported regression includes user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level. The omitted day is March 1st, 2022, the day immediately preceding the introduction of the ban. The vertical dashed line marks the date of the policy intervention.

TABLE I.2.
IMPACT OF THE BAN ON PRO-RUSSIA AND OVERALL WAR-RELATED CONTENT
EXCLUDING ACCOUNTS CREATED POST-BAN

Dependent variable	**Any Pro-RU**	**Total Tweets**	**# Pro-RU Tweets**	
Margin	**Extensive**	**Volume**	**Unrestricted**	**Cond. on Active Day**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	-0.054 (0.019) [0.006]	-0.069 (0.036) [0.058]	-0.146 (0.032) [0.000]	-0.117 (0.027) [0.000]
Unconditional	✓	✓	✓	
Conditional on Posting				✓
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.138	0.579	0.208	0.757
Approx. Percentage Change	-5.25	-6.64	-13.60	-11.04
Observations	1138236	1138236	1138236	215329

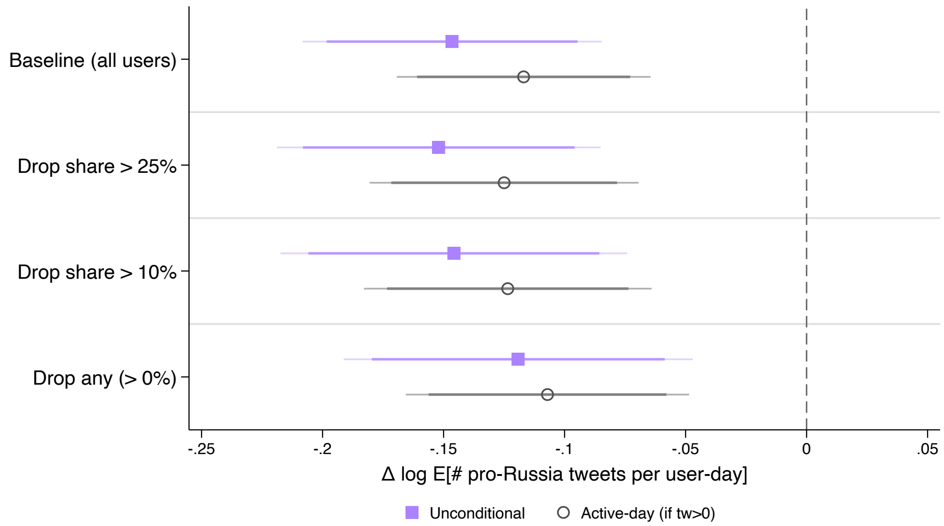
Notes: The table re-estimates the main Table II specification on the sample excluding accounts created after the ban took effect. All four columns are estimated via Poisson Pseudo-Maximum Likelihood (PPML). The coefficient of interest is the Ban × EU interaction. All specifications include user and date fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. Outcomes across columns match Table II: Column 1 is an indicator for posting any pro-Russia content on a given day (extensive margin); Column 2 is total tweet volume; Column 3 is the count of pro-Russia tweets on the balanced panel; Column 4 restricts to days with at least one tweet and captures compositional shift on active days. Percentage changes are computed as $(e^{\beta} - 1) \cdot 100$. Pre-ban means and percentage changes are computed on each column's estimation sample. Standard errors are clustered at the user level.

I.3 SUTVA: Cross-treatment status connections

In this subsection we test the sensitivity of the population-wide two-way DiD to violations of the stable unit treatment value assumption (SUTVA), which requires that the ban's effect on one user not depend on the treatment status of others. The relevant violation is communication across the ban boundary: an EU user who mentions or replies to a non-EU account (or vice versa) may still receive non-EU pro-Russia content after the ban, diluting the EU-vs-non-EU contrast and biasing the estimate toward zero. We proxy for exposure with a tweet-level indicator that flags any cross-boundary mention, aggregated to a pre-ban user-level share of a user's war-related tweets that cross the boundary; using pre-ban tweets only avoids endogeneity with the treatment. Cross-boundary communication is rare—only 7% of users have any pre-ban cross-boundary tweet—so the share is zero for over 90% of users and distribution percentiles are uninformative. We therefore exclude users above fixed share thresholds and trace the estimate across an increasingly aggressive gradient: dropping users whose cross-boundary share exceeds 25%, then 10%, and finally any user with a positive share.

The most aggressive cut—dropping any user with even a single cross-boundary tweet—is conservative and partly mechanical: a more active user is mechanically more likely to have at least one such tweet, so this exclusion disproportionately removes high-volume producers rather than isolating genuinely exposed users. Figure I.3 and Table I.3 report the Ban $\times$ EU coefficient across the gradient, for both the unconditional count of pro-Russia tweets and the active-day margin. The estimate is stable throughout: the unconditional effect moves from -0.146 at baseline to -0.152 ($> 25\%$), -0.146 ($> 10\%$), and -0.119 under the all-or-nothing cut, remaining statistically significant at every threshold ($p \leq 0.001$). The active-day margin is even flatter, from -0.117 to -0.107 (all $p < 0.001$). The baseline column reproduces the headline estimate exactly (Column 3 of Table II). The mild attenuation at the most extreme cut is consistent with its mechanical removal of active producers rather than with a spillover artifact, so we conclude that cross-boundary spillovers are unlikely to drive the population-wide effect.

FIGURE I.3.
IMPACT OF THE BAN ACROSS CROSS-NETWORK EXCLUSION THRESHOLDS (SUTVA)



Notes: The figure plots the Ban × EU coefficient from the population-wide two-way DiD (Table II), re-estimated on samples that progressively exclude users with greater pre-ban cross-network exposure, as a test for SUTVA violations arising from communication across the ban boundary. Exposure is the share of a user's pre-ban war-related tweets that mention an account on the other side of the boundary (an EU account mentioning a non-EU account, or vice versa), measured over the pre-ban window only. "Baseline" uses the full estimation sample (identical to Table II); the remaining rows drop users whose exposure share exceeds 25%, 10%, and 0% (any positive share). Because cross-boundary communication is rare—only 7% of users have any—the share is zero for over 90% of users, so we use fixed share thresholds rather than distribution percentiles. Violet squares plot the unconditional count of pro-Russia tweets per user-day (Column 3 of Table II); grey circles condition on active days (Column 4). Markers show point estimates with 90% (inner) and 95% (outer) confidence intervals. All specifications are estimated via Poisson Pseudo-Maximum Likelihood with user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level. Exact coefficients, percentage changes, and observation counts appear in Table I.3.

TABLE I.3.
IMPACT OF THE BAN ACROSS CROSS-NETWORK EXCLUSION THRESHOLDS (SUTVA)

Panel A: Unconditional (# pro-Russia tweets, balanced panel)

Users excluded	**None (baseline)**	**Share** > 25%	**Share** > 10%	**Any** (> 0%)
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	-0.146	-0.152	-0.146	-0.119
	(0.032)	(0.034)	(0.037)	(0.037)
	[0.000]	[0.000]	[0.000]	[0.001]
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.207	0.208	0.205	0.183
Approx. Percentage Change	-13.63	-14.10	-13.56	-11.23
Observations	1141514	1104004	1070322	1045264

Panel B: Active-day (# pro-Russia tweets | at least one tweet that day)

Users excluded	**None (baseline)**	**Share** > 25%	**Share** > 10%	**Any** (> 0%)
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban × EU	-0.117	-0.125	-0.123	-0.107
	(0.027)	(0.028)	(0.030)	(0.030)
	[0.000]	[0.000]	[0.000]	[0.000]
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.757	0.762	0.770	0.712
Approx. Percentage Change	-11.03	-11.75	-11.62	-10.15
Observations	215723	207559	195897	181472

Notes: Each column re-estimates the population-wide two-way DiD (Table II) on a sample that excludes users above a fixed pre-ban cross-network exposure threshold—the share of a user's pre-ban war-related tweets that cross the ban boundary: none (baseline), more than 25%, more than 10%, and any positive share. Panel A is the unconditional count of pro-Russia tweets on the balanced panel (Column 3 of Table II); Panel B conditions on active days (Column 4). All specifications are estimated via Poisson Pseudo-Maximum Likelihood with user and date fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level. Approximate percentage changes are $(e^{\beta} - 1) \cdot 100$. The baseline column reproduces the headline estimate; the coefficient is stable and remains significant across the gradient, including the most aggressive cut, which removes any user with cross-network communication and thereby also the most active producers.

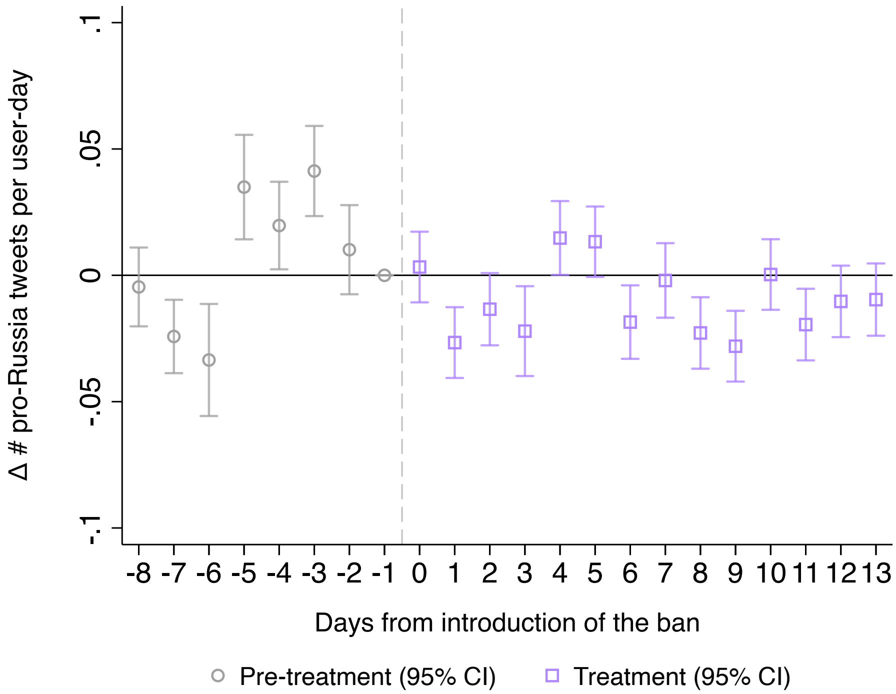
I.4 PPML vs. OLS

In this subsection, we provide additional details on our econometric choices and consider potential alternatives. Recall that our main analysis (Table II) reports four outcomes – an indicator for posting any pro-Russia content on a given day, total tweet volume, the count of pro-Russia tweets on the balanced panel, and the count of pro-Russia tweets on days with at least one tweet – all estimated by Poisson Pseudo-Maximum Likelihood (PPML).[22] The count outcomes pose challenges that motivate this estimator choice and require more careful consideration. The variable capturing the number of pro-Russia tweets posted, conditional on users posting about the war, is heavily skewed toward zero and follows a distribution that resembles a power law. Most users post no pro-Russia tweets at all, while a small minority produce many. A common approach among economists facing such distributions has been to apply a logarithmic transformation (adding one unit) and then estimate a simple OLS model. However, recent work by Chen and Roth (2024) shows that this transformation can introduce important biases and recommends alternative strategies. One such strategy is to avoid transformation altogether and instead rely on models from the Poisson family. In line with this recommendation, our main analysis employs Poisson-Pseudo Maximum Likelihood (PPML) specifications whenever we study the absolute number of pro-Russia tweets.

Although we are confident in this methodological choice, in this section we also present an alternative specification. Rather than transforming the dependent variable, we estimate OLS models directly on the raw count of pro-Russia tweets. While not inherently incorrect, this approach risks introducing distortion in the magnitude of the estimated effects. Still, it provides a useful robustness check. Figure I.4, Table I.4, and Figure I.5 reproduce the most important results of our analysis, substituting PPML models with simple OLS models. Each OLS specification is estimated on the same sample as its PPML counterpart (the units the Poisson model retains), so the comparison isolates the role of the functional form. Reassuringly, the OLS estimates reproduce the PPML findings: the ban's effects on the extensive margin, on total posting volume, and on total pro-Russia output (Columns 1–3) are all negative and statistically significant, with the extensive-margin estimate almost identical to its PPML counterpart (-5.2% versus -5.3%). As expected, the linear model attenuates the magnitudes on these skewed count outcomes—most starkly on the conditional active-day margin (Column 4), the one outcome on which the two estimators diverge, where the OLS estimate is close to zero and insignificant. The functional form thus affects the magnitude of the estimated effects but leaves the sign and significance of our main results intact.

[22] We use PPML uniformly across all four columns, including for the binary extensive-margin outcome in Column 1. PPML is consistent for binary outcomes when the conditional mean is bounded in $[0, 1]$ (Chen and Roth 2024), with the coefficient interpretable as the multiplicative change in the conditional probability via $(e^{\beta} - 1) \cdot 100$. Using a single estimator across all four columns keeps the percentage-change interpretation consistent. The OLS / linear-probability counterpart is reported in Table I.4 below as a robustness check.

FIGURE I.4.
DAILY EVENT STUDY: IMPACT OF THE BAN ON PRO-RUSSIA CONTENT IN EU VS. NON-EU



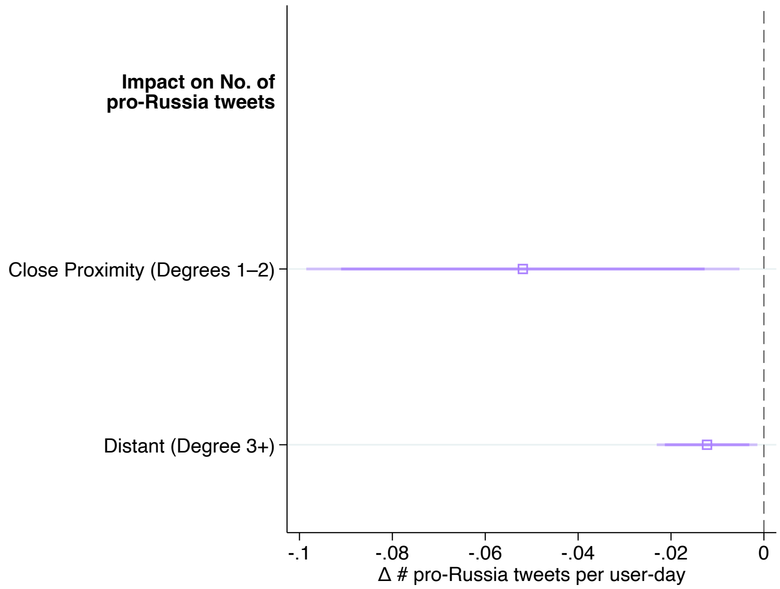
Notes: The figure displays coefficients and 95% confidence intervals from estimating an OLS event-study version of Equation 2. The regression is estimated on the same sample as the corresponding PPML event study – the user-day panel restricted to the users the Poisson model retains (those with non-zero pro-Russia output) – over February 22^{nd} to March 15^{th}, so the OLS and PPML estimates are directly comparable. The dependent variable is the number of pro-Russia tweets posted per user per day; thus, the coefficients capture the daily average effect of the ban on pro-Russia tweet activity per user-day. The reported regression includes user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level. The omitted day is March 1^{st}, 2022, the day immediately preceding the introduction of the ban. The vertical dashed line marks the date of the policy intervention.

TABLE I.4.
IMPACT OF THE BAN ON PRO-RUSSIA AND OVERALL WAR-RELATED CONTENT
OLS ESTIMATES

Dependent variable	**Any Pro-RU**	**Total Tweets**	**# Pro-RU Tweets**	
Margin	**Extensive**	**Volume**	**Unrestricted**	**Cond. on Active Day**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban $\times$ EU	-0.007	-0.046	-0.016	-0.004
	(0.001)	(0.014)	(0.004)	(0.017)
	[0.000]	[0.001]	[0.000]	[0.816]
Unconditional	✓	✓	✓	
Conditional on Posting				✓
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Pre-Ban Attitude on Russia: User	✓	✓	✓	✓
Pre-Ban Trade with Russia: State	✓	✓	✓	✓
Pre-Ban Avg. Outcome: Treated	0.137	0.577	0.207	0.757
Approx. Percentage Change	-5.17	-7.90	-7.52	-0.53
Observations	1141514	1141514	1141514	215723

Notes: The table re-estimates the main Table II specification using OLS instead of Poisson Pseudo-Maximum Likelihood, on the same estimation sample (the units the Poisson model retains, i.e. users with non-zero pro-Russia output). The coefficient of interest is the Ban $\times$ EU interaction. All specifications include user and date fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator. Outcomes across columns match Table II: Column 1 is an indicator for posting any pro-Russia content on a given day (extensive margin); Column 2 is total tweet volume; Column 3 is the count of pro-Russia tweets on the balanced panel; Column 4 restricts to days with at least one tweet and captures compositional shift on active days. Because the model is linear, percentage changes are computed as $\hat{\beta}/\bar{y}_{\text{pre}}$ rather than $(e^{\beta} - 1) \cdot 100$. Pre-ban means and percentage changes are computed on each column's estimation sample. Standard errors are clustered at the user level.

FIGURE I.5.
IMPACT OF THE BAN BY PROXIMITY TO THE OUTLETS



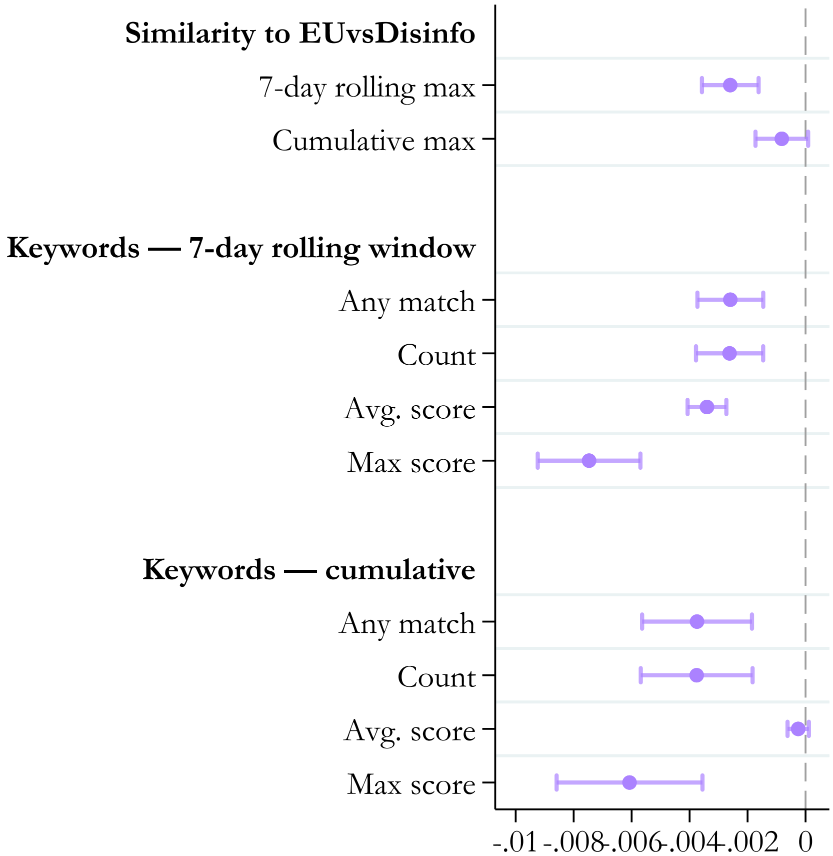
Notes: The figure displays coefficients, 90% confidence intervals (dark violet), and 95% confidence intervals (light violet) from estimating an OLS version of the static two-way DiD specification (Equation 2) for subgroups of users distinguished by their connection to the banned outlets. We present results from two separate regressions that share an identical specification but differ in the user samples analyzed. Both regressions are estimated on the PPML estimation sample (the units the Poisson model retains in Figure E.5), so the OLS and PPML by-proximity estimates are directly comparable. We include only observations between February 22nd and March 15th, to ensure a large enough sample size for each day. The first (top) coefficient shows the impact of the ban on users with high proximity to Russia Today and Sputnik, defined as users who directly retweeted or replied (first-degree connection) to posts from these outlets or interacted with posts by users who had a first-degree connection. The second coefficient reflects the impact on users with weaker connections, defined as third-degree proximity or higher. All reported regressions include user and day fixed effects and control for the user's pre-ban share of pro-Russia content and the country's pre-ban trade exposure with Russia, both interacted with the post-ban indicator; standard errors are clustered at the user level.

J Effect of the Ban on the Spread of Misinformation

To assess whether the ban reduced users' exposure to active disinformation narratives–beyond the reduction in RT/Sputnik content itself–we construct two families of measures that link tweets in our sample to cases collected and debunked by EUvsDisinfo, the EU's dedicated disinformation-tracking initiative. Our reference set consists of the 26 disinformation cases published by EUvsDisinfo between February 1 and March 31, 2022, covering the period immediately surrounding Russia's invasion of Ukraine and the subsequent broadcasting ban. The first family of measures captures *semantic similarity*. We embed each tweet and each EUvsDisinfo case (title plus summary) using the `BAAI/bge-small-en-v1.5` sentence transformer and compute cosine similarity between every tweet–case pair. We then aggregate across cases using two temporal windows: a 7-day rolling maximum (the highest similarity to any case published in the seven days preceding the tweet) and a cumulative maximum (the highest similarity to any case published up to the tweet date). The rolling window captures proximity to narratives that are currently circulating, and the cumulative measure captures exposure to any debunked narrative the user may have encountered. The second family uses *keyword matching*: each EUvsDisinfo case is characterised by three anchor keywords extracted from its title. A tweet is counted as matching a case if it contains at least two of the three anchors. We construct four variants for each temporal window–any match, count of matched cases, average match score, and maximum match score–yielding ten measures in total.

We estimate the effect of the ban on each measure separately using a simple difference-in-differences specification (EU × Ban) with user and date fixed effects. Figure J.1 plots the ten coefficients.

FIGURE J.1.
THE BAN'S EFFECT ON THE SPREAD OF MISINFORMATION



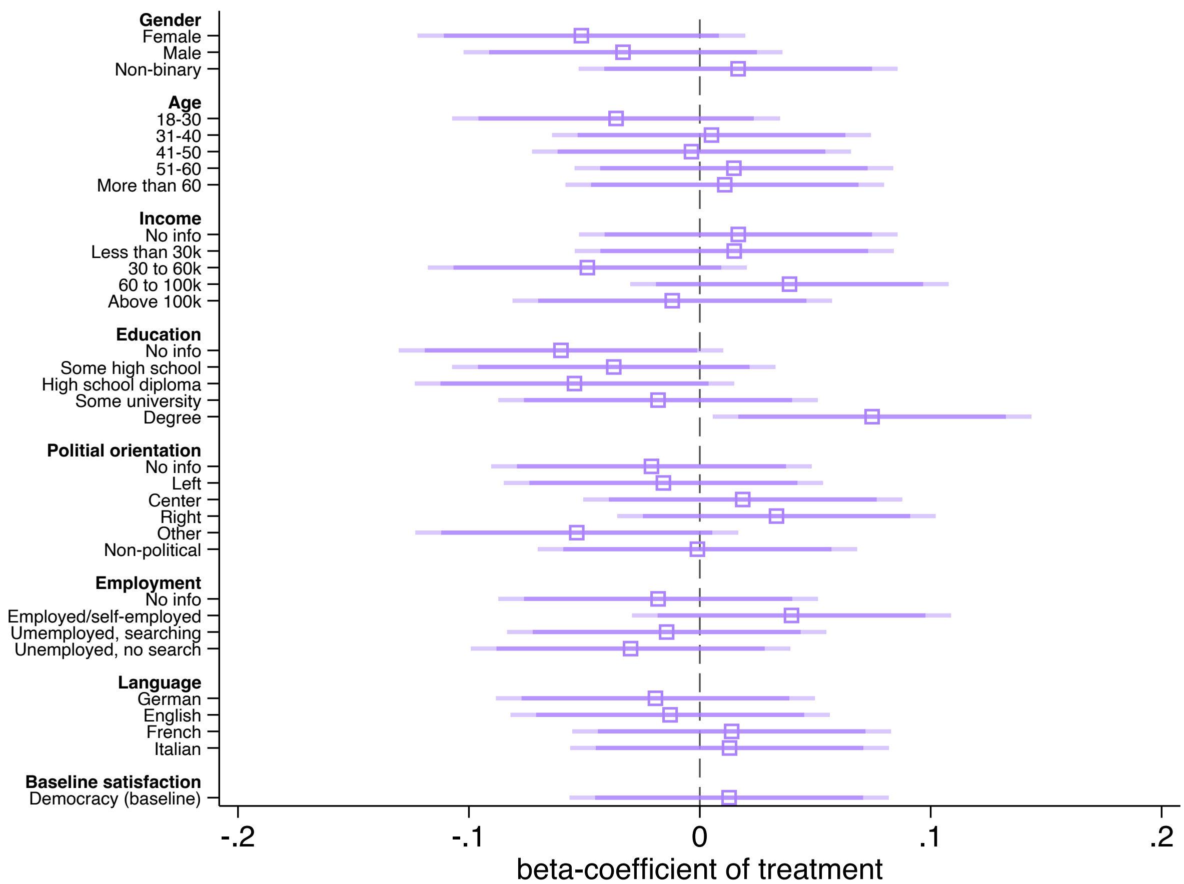
Notes: The figure displays point estimates and 95% confidence intervals for the effect of the ban on users' proximity to active disinformation narratives, estimated via a simple difference-in-differences specification. The coefficient of interest is the Ban × EU interaction. All specifications include user and date fixed effects. The sample is restricted to user-days on which the user posted at least one tweet. The reference set consists of the 26 disinformation cases published by EUvsDisinfo between February 1 and March 31, 2022. The first group of measures captures semantic similarity: each original tweet (retweets excluded) is embedded using the `BAAI/bge-small-en-v1.5` sentence transformer and its cosine similarity to each EUvsDisinfo case (headline plus summary) is computed; the 7-day rolling maximum is the highest similarity to any case published in the seven days preceding the tweet, and the cumulative maximum is the highest similarity to any case published up to the tweet date. The second and third groups use keyword matching: each EUvsDisinfo case is characterised by three anchor keywords, and a tweet is counted as matching if it contains at least two of the three anchors. For each temporal window (7-day rolling and cumulative), four variants are constructed: an indicator for any match, the count of matched cases, the average keyword hit share, and the maximum keyword hit share. Standard errors are clustered at the user level.

K Experiment

In this Appendix, we provide additional evidence from our survey experiment reported in Section 8. Figure K.1 reports balance of sociodemographic characteristics with respect to receiving treatment. Figure K.2 reports heterogeneity of the treatment effect by baseline satisfaction with democracy in EU, measured before treatment. Table K.1 reports results of the main outcomes sequentially including controls. Tables K.2 and K.3 report treatment effects on other, secondary outcomes and filler questions as a robustness check.

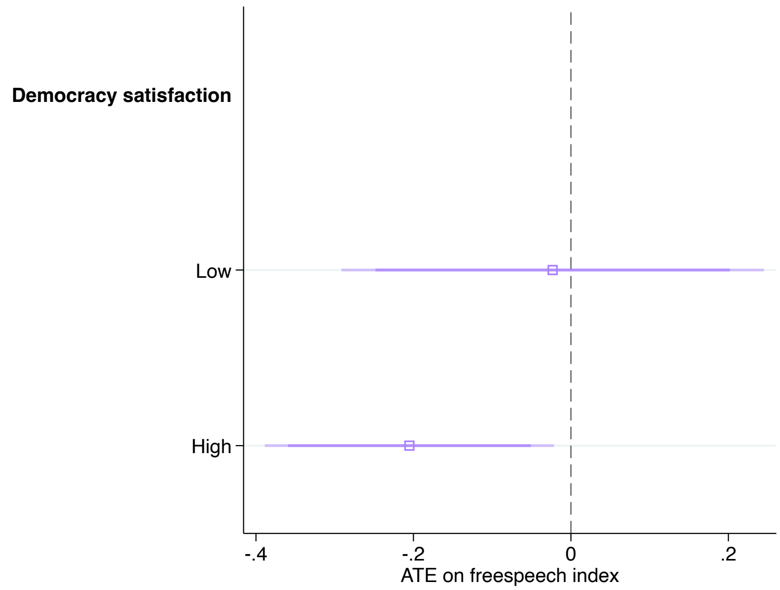
K.1 Additional Results

FIGURE K.1.
EXPERIMENT: BALANCE



Notes: The figure shows the balance of all sociodemographic characteristics with respect to treatment status of being exposed to a brief informing about the ban of Russia Today and Sputnik in our survey experiment (see Section 8 for details). We drop respondents who fail the attention check and have a duration to complete the survey below the 5th and above the 95th percentile in the duration distribution to exclude unreliable respondents.

FIGURE K.2.
EXPERIMENT: BASELINE DEMOCRACY SATISFACTION HETEROGENEITY



Notes: The figure shows the treatment effect of being exposed to a brief informing about the ban of Russia Today and Sputnik in our survey experiment (see Section 8 for details) by baseline satisfaction with the functioning of democracy in the EU of the respondent, elicited before treatment. Dependent variable is the *Freespeech Index* computed as the average of the *Freedom of speech* and *Media Independence* analysis variables. All items were elicited on a 7-point Likert approval scale and coded/rescaled for analysis so that higher values indicate stronger perceived EU protection of the corresponding construct. The media-independence item was elicited in reverse wording and aligned to this higher-is-more-protection direction. We drop respondents who fail the attention check and have a duration to complete the survey below the 5th and above the 95th percentile in the duration distribution to exclude unreliable respondents. All specifications include language fixed effects and the full set of sociodemographic controls.

TABLE K.1.
EXPERIMENT: SEQUENTIAL INCLUSION OF CONTROLS

Dependent Variable	**Freespeech index**		**Freedom of speech**		**Media independence**	
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Media Ban Information	-0.090 (0.093) [0.333]	-0.128 (0.087) [0.141]	-0.115 (0.102) [0.258]	-0.167 (0.097) [0.086]	-0.057 (0.112) [0.613]	-0.082 (0.108) [0.446]
Language FEs	✓	✓	✓	✓	✓	✓
Baseline satisfaction with democracy		✓		✓		✓
Individual characteristics						
Mean dep. var.	0.484	0.484	0.932	0.932	0.031	0.031
Observations	804	800	807	803	804	800

Notes: The table shows the treatment effect of being exposed to a brief informing about the ban of Russia Today and Sputnik in our survey experiment (see Section 8 for details). Columns 1 and 2 use the pre-registered *Freespeech Index* as the dependent variable. It is computed as the average of the *Freedom of speech* and *Media Independence* analysis variables recorded separately in columns 3–4 and 5–6, respectively. All items were elicited on a 7-point Likert approval scale and coded/rescaled for analysis so that higher values indicate stronger perceived EU protection of the corresponding construct. The media-independence item was elicited in reverse wording and aligned to this higher-is-more-protection direction. We drop respondents who fail the attention check and have a duration to complete the survey below the 5th and above the 95th percentile in the duration distribution to exclude unreliable respondents. Columns 1, 3, and 5 include language fixed effects; Columns 2, 4, and 6 additionally include baseline satisfaction with democracy elicited before treatment. Sociodemographic controls are not included in this sequential-controls table; the fully controlled specification is reported in Table VI.

TABLE K.2.
EXPERIMENT: OTHER OUTCOMES

Dependent Variable	Citizen rights	Democratic norms	Trust nat. parliament	Trust nat. gov.	Trust EU	Trust EC
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Media Ban Information	0.003	-0.019	0.021	0.012	-0.011	0.032
	(0.086)	(0.080)	(0.031)	(0.028)	(0.024)	(0.027)
	[0.968]	[0.816]	[0.501]	[0.661]	[0.646]	[0.232]
Language FEs	✓	✓	✓	✓	✓	✓
Baseline satisfaction with democracy	✓	✓	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓	✓	✓
Mean dep. var.	0.752	0.759	0.439	0.309	0.736	0.665
Observations	800	801	803	803	803	803

Notes: The table shows the treatment effect of being exposed to a brief informing about the ban of Russia Today and Sputnik in our survey experiment (see Section 8 for details). Columns 1 and 2 use items with a 7-point Likert scale of approval of whether the European Union protects the corresponding construct as the dependent variable, coded so higher values indicate stronger perceived democratic protections. Columns 3 to 6 use post-treatment binary questions about whether respondents tend to trust or not trust the corresponding institution. We drop respondents who fail the attention check and have a duration to complete the survey below the 5th and above the 95th percentile in the duration distribution to exclude unreliable respondents. All specifications include language fixed effects and the full set of sociodemographic controls, as well as the baseline satisfaction with democracy elicited before the treatment was administered.

TABLE K.3.
EXPERIMENT: FILLER QUESTIONS

Dependent Variable	Humanitarian aid	Financial support	Refugee support	Sanctions
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Media Ban Information	-0.016	0.063	0.197	0.169
	(0.075)	(0.083)	(0.085)	(0.108)
	[0.831]	[0.451]	[0.021]	[0.116]
Language FEs	✓	✓	✓	✓
Baseline satisfaction with democracy	✓	✓	✓	✓
Individual characteristics	✓	✓	✓	✓
Mean dep. var.	1.362	1.076	0.901	0.111
Observations	803	801	801	800

Notes: The table shows the treatment effect of being exposed to a brief informing about the ban of Russia Today and Sputnik in our survey experiment (see Section 8 for details) on filler questions. The filler outcomes are coded as in the analysis data and are not directly comparable across columns. We drop respondents who fail the attention check and have a duration to complete the survey below the 5th and above the 95th percentile in the duration distribution to exclude unreliable respondents. All specifications include language fixed effects and the full set of sociodemographic controls, as well as the baseline satisfaction with democracy elicited before the treatment was administered.

K.2 Questionnaire

The questionnaire appendix lists the translated question blocks as displayed in the instrument; the analysis uses the pre-treatment baseline satisfaction-with-democracy item as the covariate and treats the later trust/democracy block as post-treatment outcomes.

Censorship in Democracy: Survey

Start of Block: Default Question Block

split_language **[eng]** *Thank you for considering taking part in this survey. Which language would you prefer to use?* **[fr]** *Merci d'envisager de participer à cette enquête. Dans quelle langue préféreriez-vous poursuivre?* **[ger]** *Vielen Dank, dass Sie in Erwägung ziehen, an dieser Umfrage teilzunehmen. In welcher Sprache möchten Sie fortfahren?* **[it]** *Grazie per aver preso in considerazione la partecipazione a questa indagine. In quale lingua preferirebbe procedere?*

- ◯ English (1)
- ◯ Français (2)
- ◯ Deutsch (3)
- ◯ Italiano (4)

End of Block: Default Question Block

Start of Block: Consent Form

consent_form *Please, read carefully, thank you!* This research, conducted by researchers at the University of Bern and University of Zurich, Switzerland, is independent and aims to gather insights solely for academic purposes. The survey will take approximately **7 minutes** to complete. Compensation for participation is contingent upon completing the survey attentively. Responses flagged as insufficient may result in disqualification from payment. The survey collects personal information, including socio-demographic data. All data will be used in anonymous form. Participation is voluntary, and you may withdraw at any time. It is essential for us as researchers and for the scientific validity of this research project that you **answer the questionnaire based on your personal knowledge and opinions**. For any questions or concerns, contact Matteo Grigoletto at: matteo.grigoletto@unibe.ch Do you consent to participate?

- ◯ Yes, I consent to participate (1)
- ◯ No, I do not consent to participate (2)

End of Block: Consent Form

Page 1 of 53

Start of Block: Prolific ID

prolific_id What is your Prolific ID? **Please note that this response should auto-fill with the correct ID.**

____

End of Block: Prolific ID

Start of Block: Introduction Post-Consent

introduction *Thank you for taking part in this survey on public opinion about the EU's response to the Russia-Ukraine conflict!* ***We appreciate your time and value your opinions!*** *First, we would like you to read some brief information. Reading each text should take around 30 seconds, but this varies for each person and a* ***button to move to the next page will appear in 10 seconds****.*

End of Block: Introduction Post-Consent

Start of Block: Control: Brief 1

control_brief1 On 28 June 2024, the EU Council approved **€1 billion** in emergency aid for Ukraine through the Civil Protection Mechanism. The money will be used for **field hospitals, mobile generators, shelters and medical supplies** for people displaced ahead of the 2024-25 winter. Items will ship from the EU's RescEU stockpiles, with transport costs paid from the EU budget. About 3,000 tonnes of aid are due to arrive in the first six weeks. *Source: Council Implementing Decision (EU) 2024/1529, adopted 28 June 2024*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Control: Brief 1

Start of Block: Control: Brief 2

control_brief2 On 18 March 2025, the European Commission launched the four-year **€50 billion** assistance programme "Ukraine Facility", to keep the country's finances running during the war. The programme combines low-interest loans and grants to pay for **pensions, schools and power-line repairs**. Each payment is released only after Ukraine meets agreed anti-corruption steps and is backed by the EU budget. Fresh funds are planned every three months once the conditions are met. *Source: Regulation (EU) 2025/447 and Commission Implementing Decision C(2025) 1763*

---

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Control: Brief 2

Start of Block: Treatment: Brief 1

treat_brief1 On 28 June 2024, the EU Council approved **€1 billion** in emergency aid for Ukraine through the Civil Protection Mechanism. The money will be used for **field hospitals, mobile generators, shelters and medical supplies** for people displaced ahead of the 2024-25 winter. Items will ship from the EU's RescEU stockpiles, with transport costs paid from the EU budget. About 3,000 tonnes of aid are due to arrive in the first six weeks. *Source: Council Implementing Decision (EU) 2024/1529*

---

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Brief 1

Start of Block: Treatment: Brief 2

treat_brief2 On 18 March 2025, the European Commission launched the four-year **€50 billion** assistance programme "Ukraine Facility", to keep the country's finances running during the war. The programme combines low-interest loans and grants to pay for **pensions, schools and power-line repairs**. Each payment is released only after Ukraine meets agreed anticorruption steps and is backed by the EU budget. Fresh funds are planned every three months once the conditions are met. *Source: Regulation (EU) 2025/447 and Commission Implementing Decision C(2025) 1763*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Brief 2

Start of Block: Treatment: Ban

treat_ban On 2 March 2022 the Council of the European Union adopted a regulation that **fully suspends the broadcasting and online distribution of the Russian-state media outlets Russia Today and Sputnik** within the EU. The decision applies to television, radio, websites and social-media accounts and remains in force until the Council decides otherwise. Penalties for non-compliance include fines and withdrawal of operating licences. The measure is **binding on all member states** and became applicable immediately after its publication in the Official Journal of the EU. *Source: Council Regulation (EU) 2022/350.*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Ban

Start of Block: Attention Check

intro_attention Thank you for reading the information provided! Next, we will ask you **your opinion on the EU's response to the Russia-Ukraine conflict**. Before that, to show you have read the information carefully, please reply to the following question.

Page Break

attention_check Exclusively based on the EU-related information you just read, which of the following actions has the European Union recently taken?

- ▢ Increased the budget of the Creative Europe programme to support cinemas and theaters. (2)
- ▢ Approved €1 billion in emergency aid for Ukraine to fund field hospitals, generators, shelters, and medical supplies. (1)
- ▢ Proposed a legally binding target to cut EU greenhouse-gas emissions by at least 55% by 2030. (3)

End of Block: Attention Check

Start of Block: Primary Carousel

intro_primary Now, you will see **four statements**. For each one of them, please indicate whether you agree or disagree with the statement, in a scale between "Strongly disagree" and "Strongly Agree". After you choose your answer, the survey will automatically show the next statement. At the fourth statement, please click the button below to continue.

Page Break

primary_carousel

	Strongly Disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
The European Union protects freedom of speech. (1)	○	○	○	○	○	○	○
The EU respects democratic norms even when under pressure. (2)	○	○	○	○	○	○	○
I trust the EU to uphold citizens' fundamental rights in times of crisis. (3)	○	○	○	○	○	○	○
The EU does not guarantee the independence of the media. (4)	○	○	○	○	○	○	○

End of Block: Primary Carousel

Start of Block: Filler Carousel

intro_filler Now, you will see **four other statements**. For each one of them, please indicate whether you agree or disagree with the statement, in a scale between "Strongly disagree" and "Strongly Agree".

Page Break

filler_carousel

	Strongly Disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
The European Union fulfills its humanitarian obligations toward civilians affected by the war in Ukraine. (1)	○	○	○	○	○	○	○
I trust the EU to deliver prompt and sufficient financial assistance to Ukraine's government during the war. (2)	○	○	○	○	○	○	○
The EU effectively protects refugees and internally displaced persons fleeing the conflict. (3)	○	○	○	○	○	○	○
The EU's economic measures and sanctions provide an adequate response to Russia's aggression.	○	○	○	○	○	○	○

(4)

End of Block: Filler Carousel

Start of Block: Trust Institutions

intro_trust How much **trust** do you have in certain institutions? Now, you will see **four institutions**. For each of the them, please indicate whether you tend to trust the institution or not.

Page Break

trust_eu The European Union

- ○ Tend not to trust (1)
- ○ Tend to trust (2)

trust_parl The (National) Parliament

- ○ Tend not to trust (1)
- ○ Tend to trust (2)

trust_gov The (National) Government

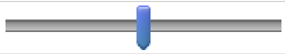
- ○ Tend not to trust (1)
- ○ Tend to trust (2)

trust_ec The European Commission

- ○ Tend not to trust (1)
- ○ Tend to trust (2)

Page Break

democracy To finish, a final question. On the whole, how satisfied are you with the way democracy works in the European Union? Please, provide your answer on a scale between 1 "Very dissatisfied" to 7 "Very satisfied".

	Very dissatisfied	Dissatisfied	Somewhat dissatisfied	Neither dissatisfied not satisfied	Somewhat satisfied	Satisfied	Very satsfied
	1	2	3	4	5	6	7
()							

End of Block: Trust Institutions

Start of Block: Personal Characteristics EU

intro_personal Now, to finish, just some questions about you!

Page Break

gender What is your gender?

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer not to say (4)

---

age What is your age bracket?

- 18-30 (1)
- 31-40 (2)
- 41-50 (3)
- 51-60 (4)
- 61-more (5)
- Prefer to not say (6)

---

income What was your total household income before taxes last year? (measure in euros)

- Less than 30,000 (1)
- 30,000 – 60,000 (2)
- 60,000 – 100,000 (3)
- Above 100,000 (4)
- Prefer to not say (5)

---

education What is the highest level of education you have completed?

- Some high school, but no diploma (1)
- High school diploma or technical institute (2)
- Some university, but no degree (3)
- Associate, bachelor, master, or graduate degree (4)
- Prefer to not say (5)

---

politics What do you consider to be your political leaning, as of today?

- Left (1)
- Right (2)
- Center (3)
- Other (4)
- Non-Political (5)
- Prefer to not say (6)

---

employment What is your current employment status?

- Employed or self-employed (1)
- Unemployed, looking for job (2)
- Unemployed, not looking (3)
- Prefer to not say (4)

End of Block: Personal Characteristics EU

Start of Block: End Message

end_message Thank you for taking part to this survey! We really appreciate it! When you proceed to the next page, you will be re-directed automatically to Prolific. See you soon!

End of Block: End Message

Start of Block: Consent Form (fr)

consent_form_fr *Veuillez lire attentivement, merci !* Cette recherche, menée par des chercheurs de l'Université de Berne et de l'Université de Zurich, en Suisse, est indépendante et vise à recueillir des informations uniquement à des fins académiques. Il faut compter environ **7 minutes** pour répondre à l'enquête. Pour être rémunéré, il est nécessaire de répondre attentivement à l'enquête. Les réponses jugées insuffisantes peuvent entraîner l'exclusion du paiement. L'enquête recueille des informations personnelles, y compris des données sociodémographiques. Toutes les données seront utilisées sous forme anonyme. La participation est volontaire et vous pouvez vous retirer à tout moment. Il est essentiel pour nous en tant que chercheurs et pour la validité scientifique de ce projet de recherche que vous **répondiez au questionnaire sur la base de vos connaissances et opinions personnelles**. Pour toute question ou préoccupation, contactez Matteo Grigoletto à l'adresse suivante : matteo.grigoletto@unibe.ch Consentez-vous à participer ?

○ Oui, je consens à participer (1)

○ Non, je ne consens pas à participer (2)

End of Block: Consent Form (fr)

Start of Block: Prolific ID (fr)

prolific_id_fr Quel est votre identifiant Prolific? **Veuillez noter que cette réponse devrait se remplir automatiquement avec l'identifiant correct.**

____

End of Block: Prolific ID (fr)

Start of Block: Introduction Post-Consent (fr)

introduction_fr *Merci d'avoir participé à cette enquête sur l'opinion publique concernant la réponse de l'UE au conflit russo-ukrainien ! Nous apprécions votre temps et vos opinions !*

Tout d'abord, nous aimerions que vous lisiez quelques brèves informations.   La lecture de chaque texte devrait prendre environ 30 secondes, mais ce délai varie d'une personne à l'autre. Un bouton permettant de passer à la page suivante apparaîtra dans 10 secondes.

End of Block: Introduction Post-Consent (fr)

Start of Block: Control: Brief 1 (fr)

control_brief1_fr Le 28 juin 2024, le Conseil de l'UE a approuvé une aide d'urgence de **€1 milliard** pour l'Ukraine par l'intermédiaire du mécanisme de protection civile. Cette somme servira à financer **des hôpitaux de campagne, des générateurs mobiles, des abris et des fournitures médicales** pour les personnes déplacées avant l'hiver 2024-25. Les articles seront expédiés à partir des stocks RescEU de l'UE, les frais de transport étant pris en charge par le budget de l'UE. Environ 3 000 tonnes d'aide devraient arriver au cours des six premières semaines.   *Source : Council Implementing Decision (EU) 2024/1529*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Control: Brief 1 (fr)

Start of Block: Control: Brief 2 (fr)

control_brief2_fr Le 18 mars 2025, la Commission européenne a lancé un programme d'aide de **€50 milliards** sur quatre ans, intitulé « Facilité pour l'Ukraine », afin de maintenir les finances du pays en état de marche pendant la guerre. Le programme combine des prêts à faible taux d'intérêt et des subventions pour financer les **retraites, les écoles et la réparation des lignes électriques**. Chaque versement n'est effectué qu'une fois que l'Ukraine a pris les mesures anticorruption convenues et qu'elle est soutenue par le budget de l'UE. De nouveaux fonds sont prévus tous les trois mois, une fois les conditions remplies.   *Source : Regulation (EU) 2025/447 and Commission Implementing Decision C(2025) 1763*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Control: Brief 2 (fr)

Start of Block: Treatment: Brief 1 (fr)

treat_brief1_fr Le 28 juin 2024, le Conseil de l'UE a approuvé une aide d'urgence de **€1 milliard** pour l'Ukraine par l'intermédiaire du mécanisme de protection civile. Cette somme servira à financer **des hôpitaux de campagne, des générateurs mobiles, des abris et des fournitures médicales** pour les personnes déplacées avant l'hiver 2024-25. Les articles seront expédiés à partir des stocks RescEU de l'UE, les frais de transport étant pris en charge par le budget de l'UE. Environ 3 000 tonnes d'aide devraient arriver au cours des six premières semaines. *Source: Council Implementing Decision (EU) 2024/1529*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Brief 1 (fr)

Start of Block: Treatment: Brief 2 (fr)

treat_brief2_fr Le 18 mars 2025, la Commission européenne a lancé un programme d'aide de **€50 milliards** sur quatre ans, intitulé « Facilité pour l'Ukraine », afin de maintenir les finances du pays en état de marche pendant la guerre. Le programme combine des prêts à faible taux d'intérêt et des subventions pour financer les **retraites, les écoles et la réparation des lignes électriques**. Chaque versement n'est effectué qu'une fois que l'Ukraine a pris les mesures anticorruption convenues et qu'elle est soutenue par le budget de l'UE. De nouveaux fonds sont prévus tous les trois mois, une fois les conditions remplies. *Source: Regulation (EU) 2025/447 and Commission Implementing Decision C(2025) 1763*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Brief 2 (fr)

Start of Block: Treatment: Ban (fr)

treat_ban_fr Le 2 mars 2022, le Conseil de l'Union européenne a adopté un règlement qui **suspend totalement la diffusion et la distribution en ligne des médias d'État russes Russia Today et Sputnik** au sein de l'UE. La décision s'applique à la télévision, à la radio, aux sites web et aux comptes de médias sociaux et reste en vigueur jusqu'à ce que le Conseil en décide autrement. En cas de non-respect de la décision, les sanctions prévues sont des amendes et le retrait des licences d'exploitation. **La mesure est contraignante pour tous les États membres** et est entrée en vigueur immédiatement après sa publication au Journal officiel de l'UE. *Source: Council Regulation (EU) 2022/350.*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Ban (fr)

Start of Block: Attention Check (fr)

intro_attention_fr Merci d’avoir pris connaissance des informations fournies ! Nous allons maintenant vous demander **votre avis sur la réponse de l’Union européenne au conflit entre la Russie et l’Ukraine**. Avant cela, afin de montrer que vous avez lu attentivement les informations, veuillez répondre à la question suivante.

Page Break

attention_check_fr En vous basant exclusivement sur les informations relatives à l'UE que vous venez de lire, quelles sont les actions suivantes que l'Union européenne a récemment entreprises ?

- ☐ Augmentation du budget du programme « Europe créative » pour soutenir les cinémas et les théâtres. (2)
- ☐ Approbation d'une aide d'urgence d'un milliard d'euros pour l'Ukraine afin de financer des hôpitaux de campagne, des générateurs, des abris et des fournitures médicales. (1)
- ☐ Proposition d'un objectif juridiquement contraignant visant à réduire les émissions de gaz à effet de serre de l'UE d'au moins 55 % d'ici à 2030. (3)

End of Block: Attention Check (fr)

Start of Block: Primary Carousel (fr)

intro_primary_fr Vous allez maintenant voir **quatre affirmations**.   Pour chacune d'entre elles, veuillez indiquer si vous êtes d'accord ou non avec l'affirmation, sur une échelle allant de « Pas du tout d'accord » à « Tout à fait d'accord ».   Une fois que vous aurez choisi votre réponse, l'enquête affichera automatiquement l'énoncé suivant. Au quatrième énoncé, veuillez cliquer sur le bouton ci-dessous pour continuer.

Page Break

primary_carousel_fr

	Pas du tout d'accord (1)	En désaccord (2)	Plutôt en désaccord (3)	Ni d'accord ni en désaccord (4)	Plutôt d'accord (5)	D'accord (6)	Tout à fait d'accord (7)
L'Union européenne protège la liberté d'expression. (1)	○	○	○	○	○	○	○
L'UE respecte les normes démocratiques, même lorsqu'elle est soumise à des pressions. (2)	○	○	○	○	○	○	○
Je fais confiance à l'UE pour défendre les droits fondamentaux des citoyens en temps de crise. (3)	○	○	○	○	○	○	○
L'UE ne garantit pas l'indépendance des médias. (4)	○	○	○	○	○	○	○

End of Block: Primary Carousel (fr)

Start of Block: Filler Carousel (fr)

intro_filler_fr Vous allez maintenant voir quatre autres affirmations.   Pour chacune d'entre elles, veuillez indiquer si vous êtes d'accord ou non avec l'affirmation, sur une échelle allant de « Pas du tout d'accord » à « Tout à fait d'accord ».

Page Break

filler_carousel_fr

	Pas du tout d'accord (1)	En désaccord (2)	Plutôt en désaccord (3)	Ni d'accord ni en désaccord (4)	Plutôt d'accord (5)	D'accord (6)	Tout à fait d'accord (7)
L'Union européenne remplit ses obligations humanitaires à l'égard des civils touchés par la guerre en Ukraine. (1)	○	○	○	○	○	○	○
Je fais confiance à l'UE pour fournir une aide financière rapide et suffisante au gouvernement ukrainien pendant la guerre. (2)	○	○	○	○	○	○	○
L'UE protège efficacement les réfugiés et les personnes déplacées à l'intérieur du pays qui fuient le conflit. (3)	○	○	○	○	○	○	○
Les mesures économiques et les sanctions de l'UE constituent une réponse adéquate à l'agression de la Russie. (4)	○	○	○	○	○	○	○

End of Block: Filler Carousel (fr)

Start of Block: Trust Institutions (fr)

intro_trust_fr Quelle **confiance** accordez-vous à certaines institutions ?   Vous allez maintenant voir **quatre institutions**. Pour chacune d'entre elles, veuillez indiquer si vous avez tendance à faire confiance à l'institution ou non.

Page Break

trust_eu_fr L'Union européenne

◯ Tendance à ne pas faire confiance (1)

◯ Tendance à faire confiance (2)

trust_parl_fr Le Parlement (national)

◯ Tendance à ne pas faire confiance (1)

◯ Tendance à faire confiance (2)

trust_gov_fr Le Gouvernement (national)

◯ Tendance à ne pas faire confiance (1)

◯ Tendance à faire confiance (2)

trust_ec_fr La Commission européenne

◯ Tendance à ne pas faire confiance (1)

◯ Tendance à faire confiance (2)

Page Break

democracy_fr Pour terminer, une dernière question. Dans l'ensemble, dans quelle mesure êtes-vous satisfait du fonctionnement de la démocratie dans l'Union européenne ? Veuillez répondre sur une échelle allant de 1 « Très insatisfait » à 7 « Très satisfait ».

	Pas du tout satisfait	Insatisfait	Plutôt insatisfait	Ni satisfait ni insatisfait	Plutôt satisfait	Satisfait	Tout à fait satisfait
	1	2	3	4	5	6	7
()							

End of Block: Trust Institutions (fr)

Start of Block: Personal Characteristics (fr)

intro_personal Pour terminer, nous avons quelques questions sur vous !

Page Break

gender_fr Quel est votre sexe ?

- ◯ Homme (1)
- ◯ Femme (2)
- ◯ Non-binaire / troisième genre (3)
- ◯ Je préfère ne pas répondre (4)

---

age_fr Quelle est votre tranche d'âge ?

- ◯ 18-30 (1)
- ◯ 31-40 (2)
- ◯ 41-50 (3)
- ◯ 51-60 (4)
- ◯ 61 ou plus (5)
- ◯ Je préfère ne pas répondre (6)

---

income_fr Quel était le revenu total de votre ménage avant impôts l'année dernière ? (mesure en euros)

- ◯ Moins de 30,000 (1)
- ◯ 30,000 – 60,000 (2)
- ◯ 60,000 – 100,000 (3)
- ◯ Plus de 100,000 (4)
- ◯ Je préfère ne pas répondre (5)

---

education_fr Quel est le niveau d'études le plus élevé que vous ayez atteint ?

- Études secondaires partielles ou moins (sans le baccalauréat) (1)
- Baccalauréat ou formation professionnelle (CAP, BEP, bac pro) (2)
- Études supérieures commencées, mais sans diplôme (3)
- BTS, DUT, licence ou diplôme supérieur (master, doctorat) (4)
- Je préfère ne pas répondre (5)

---

politics_fr Comment qualifieriez-vous votre orientation politique aujourd'hui ?

- Gauche (1)
- Droite (2)
- Centre (3)
- Autre (4)
- Apolitique (5)
- Je préfère ne pas répondre (6)

---

employment_fr Quel est votre statut professionnel actuel ?

- Employé(e) ou travailleur(se) indépendant(e) (1)
- Sans emploi, à la recherche d'un travail (2)
- Sans emploi, ne recherchant pas de travail (3)
- Je préfère ne pas répondre (4)

End of Block: Personal Characteristics (fr)

Start of Block: End Message (fr)

end_message_fr Merci d'avoir participé à cette enquête ! Nous vous en sommes vraiment reconnaissants ! Lorsque vous passerez à la page suivante, vous serez automatiquement redirigé vers Prolific.   À bientôt !

End of Block: End Message (fr)

Start of Block: Consent Form (ger)

consent_form_ger *Bitte sorgfältig lesen, danke!*   Diese von Forschern der Universität Bern und der Universität Zürich durchgeführte Studie ist unabhängig und dient ausschließlich der Gewinnung von Erkenntnissen zu wissenschaftlichen Zwecken.   Das Ausfüllen der Umfrage dauert etwa **7 Minuten**. Die Teilnahme wird nur dann vergütet, wenn Sie die Umfrage aufmerksam ausfüllen. Antworten, die als unzureichend gekennzeichnet sind, können zum Ausschluss von der Vergütung führen. In der Umfrage werden persönliche Informationen, einschließlich soziodemografischer Daten, erhoben. Alle Daten werden in anonymer Form verwendet. Die Teilnahme ist freiwillig, und Sie können sich jederzeit zurückziehen.   Für uns als Forscher und für die wissenschaftliche Validität dieses Forschungsprojekts ist es wichtig, dass Sie **den Fragebogen auf der Grundlage Ihrer persönlichen Kenntnisse und Meinungen beantworten**. Wenn Sie Fragen oder Bedenken haben, wenden Sie sich bitte an Matteo Grigoletto unter: matteo.grigoletto@unibe.ch   Sind Sie mit der Teilnahme einverstanden?

○ Ja, ich bin mit der Teilnahme einverstanden (1)

○ Nein, ich bin nicht mit der Teilnahme einverstanden (2)

End of Block: Consent Form (ger)

Start of Block: Prolific ID (ger)

prolific_id_ger Wie lautet Ihre Prolific-ID?   **Bitte beachten Sie, dass diese Antwort automatisch mit der richtigen ID ausgefüllt werden sollte.**

____

End of Block: Prolific ID (ger)

Start of Block: Introduction Post-Consent (ger) introduction_ger *Vielen Dank für Ihre Teilnahme an dieser Umfrage zur öffentlichen Meinung über die Reaktion der EU auf den Russland-Ukraine-Konflikt! Wir wissen Ihre Zeit zu schätzen und legen Wert auf Ihre Meinung! Zunächst möchten wir Sie bitten, einige kurze Informationen zu lesen. Das Lesen jedes Textes sollte etwa 30 Sekunden dauern, aber das ist von Person zu Person unterschiedlich. Nach 10 Sekunden erscheint eine Schaltfläche, mit der Sie auf die nächste Seite wechseln können.*

End of Block: Introduction Post-Consent (ger)

Start of Block: Control: Brief 1 (ger)

control_brief1_ger Am 28. Juni 2024 genehmigte der EU-Rat im Rahmen des Katastrophenschutzverfahrens eine Soforthilfe für die Ukraine in Höhe von **€1 Milliarde**. Das Geld wird für **Feldlazarette, mobile Generatoren, Notunterkünfte und medizinische Hilfsgüter** für die vor dem Winter 2024-25 vertriebenen Menschen verwendet. Die Hilfsgüter werden aus den RescEU-Lagerbeständen der EU geliefert, wobei die Transportkosten aus dem EU-Haushalt bezahlt werden. In den ersten sechs Wochen sollen etwa 3.000 Tonnen Hilfsgüter eintreffen. *Quelle: Council Implementing Decision (EU) 2024/1529*

---

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Control: Brief 1 (ger)

Start of Block: Control: Brief 2 (ger)

control_brief2_ger Am 18. März 2025 startete die Europäische Kommission das auf vier Jahre angelegte Hilfsprogramm „Ukraine-Fazilität" in Höhe von **€50 Milliarden**, um die Finanzen des Landes während des Krieges aufrechtzuerhalten. Das Programm kombiniert zinsgünstige Darlehen und **Zuschüsse zur Finanzierung von Renten, Schulen und Reparaturen von Stromleitungen**. Jede Zahlung wird erst freigegeben, wenn die Ukraine die vereinbarten Maßnahmen zur Korruptionsbekämpfung erfüllt hat und wird vom EU-Haushalt gedeckt. Sobald die Bedingungen erfüllt sind, sind alle drei Monate neue Mittel vorgesehen. *Quelle: Regulation (EU) 2025/447 and Commission Implementing Decision C(2025) 1763*

---

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Control: Brief 2 (ger)

Start of Block: Treatment: Brief 1 (ger)

treat_brief1_ger Am 28. Juni 2024 genehmigte der EU-Rat im Rahmen des Katastrophenschutzverfahrens eine Soforthilfe für die Ukraine in Höhe von **€1 Milliarde**. Das Geld wird für **Feldlazarette, mobile Generatoren, Notunterkünfte und medizinische Hilfsgüter** für die vor dem Winter 2024-25 vertriebenen Menschen verwendet. Die Hilfsgüter werden aus den RescEU-Lagerbeständen der EU geliefert, wobei die Transportkosten aus dem EU-Haushalt bezahlt werden. In den ersten sechs Wochen sollen etwa 3.000 Tonnen Hilfsgüter eintreffen.   *Quelle: Durchführungsbeschluss (EU) 2024/1529 des Rates, angenommen am 28. Juni 2024*

---

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Brief 1 (ger)

Start of Block: Treatment: Brief 2 (ger)

treat_brief2_ger Am 18. März 2025 startete die Europäische Kommission das auf vier Jahre angelegte Hilfsprogramm „Ukraine-Fazilität" in Höhe von **€50 Milliarden**, um die Finanzen des Landes während des Krieges aufrechtzuerhalten. Das Programm kombiniert zinsgünstige Darlehen und **Zuschüsse zur Finanzierung von Renten, Schulen und Reparaturen von Stromleitungen**. Jede Zahlung wird erst freigegeben, wenn die Ukraine die vereinbarten Maßnahmen zur Korruptionsbekämpfung erfüllt hat und wird vom EU-Haushalt gedeckt. Sobald die Bedingungen erfüllt sind, sind alle drei Monate neue Mittel vorgesehen.   *Quelle: Regulation (EU) 2025/447 and Commission Implementing Decision C(2025) 1763*

---

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Brief 2 (ger)

Start of Block: Treatment: Ban (ger)

treat_ban_ger Am 2. März 2022 erließ der Rat der Europäischen Union eine Verordnung, mit der die Ausstrahlung und Online-Verbreitung **vom russischen Staaat unterstützten Medienunternehmen Russia Today und Sputnik in der EU vollständig ausgesetzt wird**. Der Beschluss gilt für Fernsehen, Radio, Websites und Social-Media-Konten und bleibt in Kraft, bis der Rat anders entscheidet. Sanktionen bei Nichteinhaltung umfassen Geldstrafen und der Entzug von Betriebslizenzen. Die Maßnahme ist **für alle Mitgliedsstaaten verbindlich** und trat unmittelbar nach ihrer Veröffentlichung im Amtsblatt der EU in Kraft. *Quelle: Council Regulation (EU) 2022/350.*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Ban (ger)

Start of Block: Attention Check (ger)

intro_attention_ger Vielen Dank, dass Sie die Informationen gelesen haben! Als Nächstes werden wir Sie nach **Ihrer Meinung zur Reaktion der EU auf den Russland-Ukraine-Konflikt fragen**. Um zu zeigen, dass Sie die Informationen aufmerksam gelesen haben, beantworten Sie bitte zuvor die folgende Frage.

Page Break

attention_check_ger Ausschließlich auf Basis der EU-bezogenen Informationen, die Sie gerade gelesen haben, welche der folgenden Maßnahmen hat die Europäische Union in letzter Zeit ergriffen ?

- ☐ Hat das Budget des Programms „Kreatives Europa“ erhöht, um Kinos und Theater zu unterstützen. (2)
- ☐ Eine Soforthilfe von 1 Milliarde € für die Ukraine genehmigt, um Feldlazarette, Generatoren, Unterkünfte und medizinische Ausrüstung zu finanzieren. (1)
- ☐ Ein rechtlich verbindliches Ziel vorgeschlagen, die Treibhausgasemissionen der EU bis 2030 um mindestens 55 % zu senken. (3)

End of Block: Attention Check (ger)

Start of Block: Primary Carousel (ger)

intro_primary_ger Nun werden Sie **vier Aussagen** sehen. Bitte geben Sie für jede dieser Aussagen an, ob Sie ihr zustimmen oder nicht zustimmen, und zwar auf einer Skala zwischen „stimme überhaupt nicht zu“ und „stimme voll und ganz zu“. Nachdem Sie Ihre Antwort ausgewählt haben, zeigt die Umfrage automatisch die nächste Aussage an. Bei der vierten Aussage klicken Sie bitte auf die Schaltfläche unten, um fortzufahren.

Page Break

primary_carousel_ger

	Stimme überhaupt nicht zu (1)	Stimme nicht zu (2)	Stimme eher nicht zu (3)	Neutral (4)	Stimme eher zu (5)	Stimme zu (6)	Stimme voll und ganz zu (7)
Die Europäische Union schützt das Recht auf freie Meinungsäußerung. (1)	○	○	○	○	○	○	○
Die EU achtet die demokratischen Normen, auch wenn sie unter Druck steht. (2)	○	○	○	○	○	○	○
Ich vertraue darauf, dass die EU in Krisenzeiten die Grundrechte der Bürgerinnen und Bürger wahrt. (3)	○	○	○	○	○	○	○
Die EU garantiert nicht die Unabhängigkeit der Medien. (4)	○	○	○	○	○	○	○

End of Block: Primary Carousel (ger)

Start of Block: Filler Carousel (ger)

intro_filler_ger Nun sehen Sie **vier weitere Aussagen**.   Bitte geben Sie für jede dieser Aussagen an, wie sehr Sie ihr zustimmen oder nicht zustimmen, auf einer Skala von „Stimme überhaupt nicht zu“ bis „Stimme voll und ganz zu“.

Page Break

filler_carousel_ger

	Stimme überhaupt nicht zu (1)	Stimme nicht zu (2)	Stimme eher nicht zu (3)	Neutral (4)	Stimme eher zu (5)	Stimme zu (6)	Stimme voll und ganz zu (7)
Die Europäische Union erfüllt ihre humanitären Verpflichtungen gegenüber den vom Krieg in der Ukraine betroffenen Zivilist*innen. (1)	○	○	○	○	○	○	○
Ich vertraue darauf, dass die EU der ukrainischen Regierung während des Krieges rasch und in ausreichendem Umfang finanzielle Hilfe leistet. (2)	○	○	○	○	○	○	○
Die EU schützt Geflüchtete und Binnenvertriebene, die vor dem Konflikt fliehen, wirksam. (3)	○	○	○	○	○	○	○
Die wirtschaftlichen Maßnahmen und Sanktionen der EU stellen eine angemessene Reaktion auf die Aggression Russlands dar. (4)	○	○	○	○	○	○	○

End of Block: Filler Carousel (ger)

Start of Block: Trust Institutions (ger)

intro_trust_ger Wie viel **Vertrauen** haben Sie in bestimmte Institutionen?   Nun sehen Sie **vier Institutionen**. Bitte geben Sie für jede an, ob Sie ihr eher vertrauen oder eher nicht vertrauen.

Page Break

trust_eu_ger Europäische Union

- ◯ Eher kein Vertrauen (1)
- ◯ Eher Vertrauen (2)

---

trust_parl_ger (Nationales) Parlament

- ◯ Eher kein Vertrauen (1)
- ◯ Eher Vertrauen (2)

---

trust_gov_ger (Nationale) Regierung

- ◯ Eher kein Vertrauen (1)
- ◯ Eher Vertrauen (2)

---

trust_ec_ger Europäische Kommission

- ◯ Eher kein Vertrauen (1)
- ◯ Eher Vertrauen (2)

Page Break

democracy_ger Zum Schluss diese Abschnitts eine letzte Frage. Insgesamt betrachtet, wie zufrieden sind Sie damit, wie die Demokratie in der Europäischen Union funktioniert? Bitte beantworten Sie auf einer Skala von 1 „Sehr unzufrieden“ bis 7 „Sehr zufrieden“.

	Sehr unzufrieden	Unzufrieden	Eher unzufrieden	Weder zufrieden noch unzufrieden	Eher zufrieden	Zufrieden	Sehr zufrieden
	1	2	3	4	5	6	7
()							

End of Block: Trust Institutions (ger)

Start of Block: Personal Characteristics (ger)

intro_personal_ger Nun, zum Abschluss noch ein paar Fragen zu Ihnen!

Page Break

gender_ger Was ist Ihr Geschlecht?

- ◯ Männlich (1)
- ◯ Weiblich (2)
- ◯ Nicht-binär / Drittes Geschlecht (3)
- ◯ Möchte ich nicht angeben (4)

---

age_ger In welche Altersgruppe fallen Sie?

- ◯ 18-30 (1)
- ◯ 31-40 (2)
- ◯ 41-50 (3)
- ◯ 51-60 (4)
- ◯ 61 oder älter (5)
- ◯ Möchte ich nicht angeben (6)

---

income_ger Was war Ihr Haushaltseinkommen (vor Steuern) im vergangenen Jahr? (in Euro)

- ◯ Weniger als 30,000 (1)
- ◯ 30,000 – 60,000 (2)
- ◯ 60,000 – 100,000 (3)
- ◯ Mehr als 100,000 (4)
- ◯ Möchte ich nicht angeben (5)

---

education_ger Welchen höchsten Bildungsabschluss haben Sie erreicht?

- ◯ Kein Schulabschluss / höchstens einige Schuljahre (1)
- ◯ Schulabschluss (Haupt-, Realschule, Abitur) oder berufliche Ausbildung (z. B. Lehre, Berufsfachschule) (2)
- ◯ Einige Semester Hochschule, aber kein Abschluss (3)
- ◯ Hochschul- oder Fortbildungsabschluss (Associate / Meister / Techniker / Bachelor / Master / Promotion) (4)
- ◯ Möchte ich nicht angeben (5)

---

politics_ger Wie würden Sie Ihre politische Ausrichtung heute einschätzen?

- ◯ Links (1)
- ◯ Rechts (2)
- ◯ Mitte (3)
- ◯ Andere (4)
- ◯ Unpolitisch (5)
- ◯ Möchte ich nicht angeben (6)

---

employment_ger Wie ist Ihre derzeitige Beschäftigungssituation?

- ◯ Erwerbstätig oder selbstständig (1)
- ◯ Arbeitslos, auf Jobsuche (2)
- ◯ Arbeitslos, nicht auf Jobsuche (3)
- ◯ Möchte ich nicht angeben (4)

End of Block: Personal Characteristics (ger)

Start of Block: End Message (ger)

end_message_ger Vielen Dank, dass Sie an dieser Umfrage teilgenommen haben! Wir wissen das sehr zu schätzen! Wenn Sie zur nächsten Seite weitergehen, werden Sie automatisch zu Prolific weitergeleitet.   Bis bald!

End of Block: End Message (ger)

Start of Block: Consent Form (it)

consent_form_it *La preghiamo di leggere attentamente, grazie!*   Questa ricerca, condotta da ricercatori dell'Università di Berna e dell'Università di Zurigo, in Svizzera, è indipendente e mira a raccogliere informazioni esclusivamente per scopi accademici.   Il sondaggio richiede circa **7 minuti** per essere completato. Il compenso per la partecipazione è subordinato al completamento consono del sondaggio. Le risposte segnalate come insufficienti possono comportare l'esclusione dal pagamento. Il sondaggio raccoglie informazioni personali, compresi i dati socio-demografici. Tutti i dati saranno utilizzati in forma anonima. La partecipazione è volontaria e ci si può ritirare in qualsiasi momento.   È essenziale per noi ricercatori e per la validità scientifica di questo progetto di ricerca che lei **risponda al questionario sulla base delle sue conoscenze e opinioni personali**. Per qualsiasi domanda o dubbio, contattare Matteo Grigoletto all'indirizzo: matteo.grigoletto@unibe.ch.   Acconsente a partecipare?

◯ Sì, acconsento a partecipare (1)

◯ No, non acconsento a partecipare (2)

End of Block: Consent Form (it)

Start of Block: Prolific ID (it)

prolific_id_it Qual è la sua ID Prolific?   **Cortesemente noti, la risposta dovrebbe auto-compilarsi con la corretta ID.**

____

End of Block: Prolific ID (it)

Start of Block: Introduction Post-Consent (it) introduction_it Grazie per la sua partecipazione a questo sondaggio sull'opinione pubblica in merito alla risposta dell'UE al conflitto tra Russia e Ucraina!   Apprezziamo il suo tempo e le sue opinioni!   Per prima cosa, vorremmo che leggesse alcune brevi informazioni.   La lettura di ogni testo dovrebbe durare circa 30 secondi, ma varia da persona a persona e un pulsante per passare alla pagina successiva apparirà entro 10 secondi.

End of Block: Introduction Post-Consent (it)

Start of Block: Control: Brief 1 (it)

control_breif1_it Il 28 giugno 2024, il Consiglio dell'UE ha approvato un aiuto di emergenza di **€1 miliardo** per l'Ucraina attraverso il Meccanismo di protezione civile. Il denaro sarà utilizzato per **ospedali da campo, generatori mobili, rifugi e forniture mediche** destinate alle persone sfollate in vista dell'inverno 2024-25. I beni saranno spediti dalle scorte rescEU dell'UE, con i costi di trasporto coperti dal bilancio dell'Unione. Circa 3 000 tonnellate di aiuti dovrebbero arrivare nelle prime sei settimane.   *Fonte: Council Implementing Decision (EU) 2024/1529*

---

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Control: Brief 1 (it)

Start of Block: Control: Brief 2 (it)

control_brief2_it Il 18 marzo 2025, la Commissione europea ha avviato il programma di assistenza quadriennale da **€50 miliardi** «Ukraine Facility», per mantenere in funzione le finanze del Paese durante la guerra. Il programma combina prestiti a basso interesse e sovvenzioni per pagare **pensioni, scuole e riparazioni delle linee elettriche**. Ogni erogazione viene rilasciata solo dopo che l'Ucraina soddisfa le misure anticorruzione concordate ed è garantita dal bilancio dell'UE. Nuovi fondi sono previsti ogni tre mesi, una volta soddisfatte le condizioni.   *Fonte: Regulation (EU) 2025/447 and Commission Implementing Decision C(2025) 1763*

---

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Control: Brief 2 (it)

Start of Block: Treatment: Brief 1 (it)

treat_brief1_it Il 28 giugno 2024, il Consiglio dell'UE ha approvato un aiuto di emergenza di **€1 miliardo** per l'Ucraina attraverso il Meccanismo di protezione civile. Il denaro sarà utilizzato per **ospedali da campo, generatori mobili, rifugi e forniture mediche** destinate alle persone sfollate in vista dell'inverno 2024-25. I beni saranno spediti dalle scorte rescEU dell'UE, con i costi di trasporto coperti dal bilancio dell'Unione. Circa 3 000 tonnellate di aiuti dovrebbero arrivare nelle prime sei settimane. *Fonte: Council Implementing Decision (EU) 2024/1529, adopted 28 June 2024*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Brief 1 (it)

Start of Block: Treatment: Brief 2 (it)

treat_brief2_it Il 18 marzo 2025, la Commissione europea ha avviato il programma di assistenza quadriennale da **€50 miliardi** «Ukraine Facility», per mantenere in funzione le finanze del Paese durante la guerra. Il programma combina prestiti a basso interesse e sovvenzioni per pagare **pensioni, scuole e riparazioni delle linee elettriche**. Ogni erogazione viene rilasciata solo dopo che l'Ucraina soddisfa le misure anticorruzione concordate ed è garantita dal bilancio dell'UE. Nuovi fondi sono previsti ogni tre mesi, una volta soddisfatte le condizioni.
Fonte: Regulation (EU) 2025/447 and Commission Implementing Decision C(2025) 1763

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Brief 2 (it)

Start of Block: Treatment: Ban (it)

treat_ban_it Il 2 marzo 2022 il Consiglio dell'Unione Europea ha adottato un regolamento che **sospende completamente la trasmissione e la distribuzione online delle testate giornalistiche russe Russia Today e Sputnik** all'interno dell'UE. La decisione si applica a televisione, radio, siti web e account sui social media e rimane in vigore finché il Consiglio non deciderà altrimenti. Le sanzioni per la mancata osservanza includono multe e revoca delle licenze operative. La misura è **vincolante per tutti gli Stati membri** ed è entrata in vigore immediatamente dopo la sua pubblicazione nella Gazzetta ufficiale dell'UE. *Fonte: Council Regulation (EU) 2022/350.*

timer Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Treatment: Ban (it)

Start of Block: Attention Check (it)

intro_attention_it Grazie per aver letto le informazioni fornite! Ora le chiediamo la **sua opinione sulla risposta dell'UE al conflitto tra Russia e Ucraina**. Prima di ciò, per dimostrare che ha letto attentamente le informazioni, la preghiamo di rispondere alla seguente domanda.

Page Break

attention_check_it Esclusivamente in base alle informazioni relative all'UE che ha appena letto, quale delle seguenti azioni è stata recentemente intrapresa dall'Unione europea?

- ▢ Ha aumentato il budget del programma Europa Creativa per sostenere cinema e teatri. (2)
- ▢ Ha approvato €1 miliardo di aiuti di emergenza per l'Ucraina per finanziare ospedali da campo, generatori, rifugi e forniture mediche. (1)
- ▢ Ha proposto un obiettivo legalmente vincolante di ridurre le emissioni di gas a effetto serra dell'UE di almeno il 55 % entro il 2030. (3)

End of Block: Attention Check (it)

Start of Block: Primary Carousel (it)

intro_primary_it Ora vedrà **quattro affermazioni**. Per ciascuna di esse, indichi se è d'accordo o in disaccordo con l'affermazione, su una scala da «Assolutamente in disaccordo» a «Assolutamente d'accordo». Dopo aver scelto la sua risposta, il sondaggio mostrerà automaticamente l'affermazione successiva. Alla quarta affermazione, clicchi il pulsante in basso per continuare.

Page Break

primary_carousel_it

	Assolutamente in disaccordo (1)	In disaccordo (2)	Piuttosto in disaccordo (3)	Nè d'accordo nè in disaccordo (4)	Piuttosto d'accordo (5)	D'accordo (6)	Assolutamen d'accordo (
L'Unione europea tutela la libertà di parola. (1)	○	○	○	○	○	○	○
L'UE rispetta le norme democratiche anche quando è sotto pressione. (2)	○	○	○	○	○	○	○
Ho fiducia che l'UE salvaguardi i diritti fondamentali dei cittadini in tempi di crisi. (3)	○	○	○	○	○	○	○
L'UE non garantisce l'indipendenza dei media. (4)	○	○	○	○	○	○	○

End of Block: Primary Carousel (it)

Start of Block: Filler Carousel (it)

intro_filler Ora vedrai **altre quattro affermazioni**. Per ciascuna di esse, indichi se è d'accordo o in disaccordo con l'affermazione, su una scala da «Assolutamente in disaccordo» a «Assolutamente d'accordo».

Page Break

filler_carousel_it

	Assolutamente in disaccordo (1)	In disaccordo (2)	Piuttosto in disaccordo (3)	Nè daccordo nè in disaccordo (4)	Piuttosto d'accordo (5)	D'accordo (6)	Assolutame d'accordo (
L'Unione Europea adempie ai propri obblighi umanitari nei confronti dei civili colpiti dalla guerra in Ucraina. (1)	○	○	○	○	○	○	○
Ho fiducia che l'UE eroghi un'assistenza finanziaria rapida e sufficiente al governo ucraino durante la guerra. (2)	○	○	○	○	○	○	○
L'UE protegge efficacemente i rifugiati e gli sfollati interni che fuggono dal conflitto. (3)	○	○	○	○	○	○	○
Le misure economiche e le sanzioni dell'UE costituiscono una risposta adeguata all'aggressione della Russia. (4)	○	○	○	○	○	○	○

End of Block: Filler Carousel (it)

Start of Block: Trust Institutions (it)

intro_trust_it Quanta **fiducia** ha in alcune istituzioni?   Ora vedrà **quattro istituzioni**. Per ciascuna di esse, indichi se tende a fidarsi dell'istituzione o meno.

Page Break

trust_eu_it L'Unione Europea

- ○ Tendo a non fidarmi (1)
- ○ Tendo a fidarmi (2)

---

trust_parl_it Il Parlamento (Nazionale)

- ○ Tendo a non fidarmi (1)
- ○ Tendo a fidarmi (2)

---

trust_gov_it Il Governo (Nazionale)

- ○ Tendo a non fidarmi (1)
- ○ Tendo a fidarmi (2)

---

trust_ec_it La Commissions Europea

- ○ Tendo a non fidarmi (1)
- ○ Tendo a fidarmi (2)

Page Break

democracy_it Per finire questa sezione, un'ultima domanda.   Nel complesso, quanto è soddisfatto/a del funzionamento della democrazia nell'Unione Europea?   La preghiamo di rispondere su una scala da 1 «Molto insoddisfatto» a 7 «Molto soddisfatto».

Molto insoddisfatto/a	Insoddisfatto/a	Piuttosto insoddisfatto/a	Nè soddisfatto/a nè insoddisfatto/a	Piuttosto soddisfatto/a	Soddisfatto/a	Molto soddisfatto/a
1	2	3	4	5	6	7

()

End of Block: Trust Institutions (it)

Start of Block: Personal Characteristics (it)

intro_personal_it Ora, per finire, alcune domande su di lei!

Page Break

gender_it Qual è la sua identità di genere?

- Maschio (1)
- Femmina (2)
- Non-binario / terzo genere (3)
- Preferisco non rispondere (4)

---

age_it Qual è la sua fascia d'età?

- 18-30 (1)
- 31-40 (2)
- 41-50 (3)
- 51-60 (4)
- 61 o di più (5)
- Preferisco non rispondere (6)

---

income_it Qual è stato il reddito familiare totale, al lordo delle imposte, lo scorso anno? (in euro)

- Meno di 30,000 (1)
- 30,000 – 60,000 (2)
- 60,000 – 100,000 (3)
- Più di 100,000 (4)
- Preferisco non rispondere (5)

---

education_it Qual è il livello di istruzione più alto che ha completato?

- ○ Alcuni anni di scuola superiore o meno (1)
- ○ Diploma di scuola superiore o istituto tecnico (2)
- ○ Alcuni anni di università, ma senza laurea (3)
- ○ Diploma universitario breve / laurea triennale / laurea magistrale (o titolo post-laurea) (4)
- ○ Preferisco non rispondere (5)

---

politics_it Come definirebbe il suo orientamento politico oggi?

- ○ Sinistra (1)
- ○ Destra (2)
- ○ Centro (3)
- ○ Altro (4)
- ○ Non politico/a (5)
- ○ Preferisco non rispondere (6)

---

employment_it Qual è la sua attuale situazione occupazionale?

- ○ Occupato/a o lavoratore/rice autonomo/a (1)
- ○ Disoccupato/a, in cerca di lavoro (2)
- ○ Disoccupato/a, non in cerca di lavoro (3)
- ○ Preferisco non rispondere (4)

End of Block: Personal Characteristics (it)

Start of Block: End Message (it)

end_message_it Grazie per aver partecipato a questo sondaggio! Lo apprezziamo davvero! Quando passerà alla pagina successiva, verrà reindirizzato/a automaticamente a Prolific.   A presto!

End of Block: End Message (it)