Censorship in Democracy

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

September 29, 2025

Abstract

The spread of propaganda and misinformation from autocratic regimes is a growing concern in democracies. We study the European Union's ban on Russian state-backed news outlets after the 2022 invasion of Ukraine, analyzing 677,780 tweets from 146,633 users with a difference-in-differences design in a daily panel. The ban reduced pro-Russian tweets by 10.9% per active user day, with strongest effects among users directly connected to the banned outlets. We find no evidence of substitution to secondary suppliers. Evidence on mechanisms indicates that the ban curtailed pro-Russian content by removing key agenda-setters. Finally, we examine the costs of censorship in a democratic context: A pre-registered experiment finds reduced satisfaction with free speech, particularly among political centrists.

Keywords: Censorship, Propaganda, Policy effectiveness, 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, Elliot Ash, Milena Djourelova, Ruben Durante, Ruben Enikolopov, Kai Gehring, Gaël Le Mens, Pierre-Guillaum Meon, 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, 5th Monash-Warwick-Zurich Text-as-Data workshop, Annual Meeting of the Austrian Economic Association (NOeG) 2023, Digital Democracy Workshop DigDemLab Zurich, 17th CESifo Workshop on Political Economy, 2nd CESifo/ifo Junior Workshop on Big Data, 7th Conference on Political Economy of Democracy and Dictatorship, 17th RGS Doctoral Conference, European Public Choice Society 2024, FRIAS conference Understanding the Rise of Autocrats in the 21st 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.

[*] 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 censorship of media to suppress dissent and manage the flow of information within their borders, while employing slanted narratives to exert influence abroad (Guriev and Treisman 2022). Concerns regarding the spread of misinformation, foreign propaganda, and biased narratives, especially on social media, have been rising in recent years in democratic societies. With the ever-increasing number of people who rely on social media as a primary source of information[1] and the growing evidence that misinformation and biased narratives can cause distorted political outcomes, foster affective polarization, societal upheaval, and hate crimes (Lazer et al. 2018; Allcott et al. 2020; Bursztyn, Egorov, and Fiorin 2020; Guriev, Melnikov, and Zhuravskaya 2021; Levy 2021; Müller and Schwarz 2022; Skafle et al. 2022; Müller and Schwarz 2023), the quest for effective policy solutions has received heightened attention (Persily and Tucker 2020).

Recent work by Guriev et al. (2023) highlights two major policy alternatives democracies could adopt to counter these, perceived or realized, threats. One strategy addresses the issue at the individual level with media literacy campaigns, fact-checking tools, behavioral interventions, and similar measures. These interventions have been rigorously evaluated (see, e.g. Pennycook et al. 2020; Fazio 2020; Yaqub et al. 2020; Henry, Zhuravskaya, and Guriev 2022; Arechar et al. 2023; a survey by Nyhan 2020; a toolbox by Kozyreva et al. 2024; and a meta-analysis by Pennycook and Rand 2022) and found to be capable of reducing the circulation of false news on social media. Guriev et al. (2023) go beyond the reduced form evidence of these studies and provide a unified framework for assessing the effectiveness of different small-scale policies on the circulation of both accurate and false news.

Another strategy relies on top-down legislative action to regulate content on social media platforms. Recent actions, such as the Protecting Americans from Foreign Adversary Controlled Applications Act passed by the US Congress in 2024, the EU's Digital Services Act (European Union 2023), the German NetzDG (Jiménez Durán, Müller, and Schwarz 2025) or Israel's ban on Al Jazeera's broadcasting activity, demonstrate that democratic governments see large-scale policy interventions as a viable tool to counteract the spread of misinformation. However, we lack systematic evidence on the effectiveness of the legislative, top-down approach in a democratic context.[2]

At the same time, free speech and independence of media are considered backbones of the democratic order itself.[3] Using censorship in a democratic context might undermine trust in these core features of

[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] There is some evidence on the effects of censorship in authoritarian regimes (Chen and Yang 2019; Becker, Pino, and Vidal-Robert 2021). However, systematic evidence on the effectiveness and consequences of government-imposed censorship in a democratic context is limited. Andres and Slivko (2021) and Jiménez Durán, Müller, and Schwarz (2025) show that the German NetzDG effectively reduced the spread of hateful posts online.

[3] 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 feature of "open societies".

democratic institutions themselves. It is therefore crucial to understand the effectiveness and consequences of censorship in the context of liberal democracies. Can censorship be effective in curbing the spread of misinformation and foreign propaganda? Does the use of censorship in a democratic context undermine trust in core democratic principles?

To shed light on the effects of censorship in democracies, we study the European Union's ban on Russian state-backed media outlets implemented in March 2022 in the context of Russia's invasion of Ukraine. The ban aimed to counteract the spread of Russian narratives, provided by seemingly independent news sources and used to justify the invasion and influence public opinion.[4] The unprecedented decision to completely ban all activity of the two most prominent of these outlets – Russia Today and Sputnik – was taken virtually overnight. It affected all their channels, including online platforms, in the European Union from the 2nd of March onward. We focus on the social media platform X (previously known and hereafter referred to as Twitter), a pivotal platform in shaping public opinion and fueling both offline and online political activity (Allcott and Gentzkow 2017; Acemoglu, Hassan, and Tahoun 2018). Twitter serves as a platform where narratives – often led by misinformation or radical, influential users – can become extreme (Müller and Schwarz 2023). Its influence extends beyond digital boundaries, with narratives and stories that gain traction on the platform frequently making their way into mainstream and traditional media (Cage, Herve, and Mazoyer 2022).

To measure pro-Russian content on Twitter, we leverage recent advances in AI to classify tweet content with a large language model (LLM). Our pipeline classifies each tweet as pro-Russia, neutral, or anti-Russia and assigns up to five topics identified by the model through an unsupervised procedure. We adopt this three-category scheme to simplify labeling and reduce ambiguity for the LLM: by posing contrasting, logically separable labels (pro vs. anti, with a neutral fallback), we minimize inter-class overlap and encourage clear decision boundaries. In the analysis, we focus primarily on tweets labeled pro-Russia and on the topics associated with them.

We validate the reliability of the LLM classification pipeline by constructing an alternative measure for pro-Russian content, expanding upon the approaches by Gennaro and Ash (2023) and Gentzkow and Shapiro (2011). We assemble a pro-Russian and a pro-Ukrainian reference pole by embedding and averaging across tweets from government representatives of the corresponding nations in the conflict. We then embed all tweets in our sample and compute their cosine-similarity relative to these poles. The ratio between similarity to the Russian pole relative to the Ukrainian pole then delivers an alternative, purely data-driven, measure of pro-Russian content. This alternative procedure aligns closely with the LLM classification, and all results of our core analysis remain qualitatively unchanged.

Next, we use the ban as a natural experiment that allows us to analyze the causal effect of censorship in a democratic context. We leverage the fact that during the period of our study, only the European Union – our treatment group – implemented the ban, whereas no such measure was taken in non-EU countries – our control group. We use this variation in a standard difference-in-difference estimation strategy, building on a panel on the user-day level. We document the main effect of the ban along three dimensions. First, we show that on average, the ban was effective in reducing the pro-Russian content

[4] As highlighted by the European Court of Justice's upholding, the ban was implemented on the grounds of the prohibition of "propaganda for war" in international law (International Covenant on Civil and Political Rights, UN 1966).

per active user-day by around 10.9%. Second, we go beyond the total number of tweets, and also examine whether users post at all. The ban reduced the probability of posting or sharing at least one tweet of pro-Russian content per active user-day in the EU relative to non-EU countries by around 8.5%. Third, the ban also led to a slight reduction in the probability of users tweeting or retweeting about the war in general by around 4.5%. Crucially, in an event study specification, we observe no discernible pre-trends in our outcome of interest, the number of pro-Russia tweets ("own tweets", replies, and retweets) per active user-day.

In the next step, we dig deeper into the heterogeneous impact of the ban. First, we document that the policy affects users directly connected to the outlets before the ban the most. With increasing distance to the outlets in the Twitter network or for users unconnected to the banned outlets, the effect is still detectable but clearly muted. Second, we find that the ban is particularly effective among highly active suppliers of pro-Russian content. Third, we provide evidence that users who are both highly active and highly popular – defined as those with high engagement in terms of retweets, replies, and likes by other users – reduced their self-created pro-Russia content. This evidence suggests that the ban was not only effective in influencing the spread of pro-Russian content among users directly connected to the outlets but also had a more profound impact on the overall pattern of production and consumption of pro-Russian content on Twitter.

To investigate the adjustment in supply and demand of pro-Russian content in more detail, we focus on the engagement (retweets, likes, and replies) that pro-Russian content receives. We document that, by and large, most groups of suppliers of pro-Russian content did not experience meaningful changes in engagement after the ban. However, we find suggestive evidence of a decrease in per-tweet engagement received by the highly active and highly popular pre-ban suppliers of pro-Russia content. Users in this category in the EU experience a 34% drop in the number of retweets, likes, and replies for their pro-Russian original tweets ("own tweets" and replies) relative to their counterparts not affected by the ban. This group of secondary suppliers had the closest connections to Russia Today and Sputnik before the ban, suggesting they were most dependent on the outlets for their own content. Our finding that only this group of suppliers reduced their original pro-Russia content and received less engagement after the ban is consistent with an increase in production costs for pro-Russia content once key providers of narratives are removed. We find no evidence of attempts to substitute the content of the banned outlets via increased pro-Russia tweet production among these active suppliers of pro-Russian content.

Finally, we study the mechanism behind the documented effect of the ban. We focus on the role of the outlets as key agenda setters and provide evidence that removing them from the network is associated with a lack of inspiration and guidance for other users who want to produce pro-Russian content. First, we find that pro-Russian content in the EU after the ban uses a lower number of topics per pro-Russian tweet, indicating that pro-Russian content suppliers lack inspiration to create narratives. Second, we document that the overlap between topics used by users on a given day and the topics pushed by the banned outlets in the EU after the ban, relative to their counterparts in non-EU countries that can still rely on the agenda set by Russia Today and Sputnik each day, is clearly reduced. Taken together, we take this as evidence that the banned outlets are crucial agenda setters, providing other users with cheap and easily accessible frames for narratives that they use in their production of pro-Russian content. Once the ban removes the outlets, other users face a higher cost of finding topics to use in producing and sharing pro-Russian content. This increase in costs is consistent with the overall reduction in tweets about the war, the lower probability of using a pro-Russian slant when tweeting, and the reduced total amount of pro-Russian content shared on active user-days.

Our empirical analysis of the ban on Russia Today and Sputnik in the EU finds that censorship can be an effective tool for curbing misinformation and foreign propaganda. Yet, using restrictive measures to control speech and media may risk eroding the very norms that form the foundation of the democratic order (Popper 1945; Linz 1978). To shed light on the cost of using censorship in a democratic context, we conduct an online survey experiment with 900 participants located in the European Union. We randomly assign half of the participants to receive information about the 2022 ban on all broadcasting activities of Russia Today and Sputnik – information that either introduces new knowledge for previously unaware respondents or provides a memory cue about the ban. We find that respondents who received exposure to information about the ban reduce their belief that the European Union upholds the principles of freedom of speech and independence of media (pooled together in an index) by around 0.1 of a standard deviation or around 11.3% of the pre-ban level of approval (statistically significant at the 10% level). We also detect meaningful heterogeneity with respect to the political orientation of respondents. Self-declared centrists respond negatively, while respondents who position themselves on the left or right do not display a negative reaction to the treatment. In sum, we interpret this as evidence that using censorship in a democratic context can have non-negligible costs, particularly by eroding trust in core features of the democratic order among citizens in the center of the political spectrum - arguably the core constituency upholding the democratic order. In sum, our findings highlight a dynamic tradeoff when using censorship in a democratic context. On the one hand, it may be an effective policy tool in the short run, curbing the spread of misinformation and foreign propaganda. On the other hand, the use of censorship might erode the very norms that are constituent parts of the democratic order itself in the long run.

Related Literature: This 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. Consumers may distance themselves from news 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-Robert (2021) and Blasutto and de la Croix (2023) demonstrate 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 the context of 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 contexts, Bjørnskov and Voigt (2021) study the effect of constitutional provisions about preventing media censorship in the aftermath of terrorist attacks. They provide one of the few examples that investigate the interaction between functioning constitutional systems and censorship. Furthermore, this is in line with the work by Kellam and Stein (2016), who find evidence that strong presidents can be threatening to media freedom even in democratic contexts.

A growing strand of this literature examines subtler, state- or platform-led rules that shape information flows. Information withholding that is only targeting certain outlets is referred to as "selective censorship" (Guriev and Treisman 2022). Empirical analyses of this form of censorship have been rare at this point. 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. In sum, empirical evidence on the effectiveness and costs of censorship, particularly in a democratic context, is scarce. To the best of our knowledge, our work is the first to explore a natural experiment setting to investigate the effect of censorship in democracy. The context in which we address this issue allows us to draw conclusions about a short-term, high-stakes institutional reaction, unlike past work, which has largely focused on long-term effects of censorship (Becker, Pino, and Vidal-Robert 2021). Additionally, we provide evidence of the potential dynamic adjustments of the media market to censorship in such a context. Finally, we shed light on the cost associated with the use of censorship in a democratic context. We conduct an online survey experiment directly linking our empirical setting to provide the, to our knowledge, first causal evidence of how the use of censorship can erode democratic norms.

Second, we also add to the rich literature investigating the political economy of social media (see Campante, Durante, and Tesei (2023) for an overview). 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).

There are multiple ways democratic institutions and business actors operating within democracies can address the issue of misinformation, propaganda, and biased narratives in social media. First, directly targeting the users on this platform is one way. These interventions have been rigorously evaluated (see, e.g., Pennycook et al. (2020); Fazio (2020); Yaqub et al. (2020); Henry, Zhuravskaya, and Guriev (2022); Ershov and Morales (2024); Arechar et al. (2023), a survey by Nyhan (2020), a toolbox by Kozyreva et al. (2024) and a meta-analysis by Pennycook and Rand (2022)) and found to be capable of reducing the circulation of false news on social media. Second, targeting specific accounts or content directly is an alternative policy option. The ban on Russia Today and Sputnik that we investigate in this paper can be classified as such a top-down intervention, on which there is some emerging evidence. 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 are studies by Müller and Schwarz (2022) and Jiménez Durán, Müller, and Schwarz (2025). They study the effect of banning Trump's account on reducing toxicity among his followers and the effects of a German regulation on removing online hate speech directed towards refugees. In line with our results, both studies show that online content moderation can curb toxicity and hate speech online. We contribute and go beyond the existing evidence on three dimensions. First, we can leverage a large-scale top-down intervention, censoring key media outlets directly, in contrast to small-scale interventions, which target users directly. Second, our study exploits cross-country variation in who is affected by the ban. This variation yields a more natural control group relative to existing evidence on content moderation that relies on synthetically constructed comparison groups. Third, we study the response of particular users – the pre-ban suppliers of pro-Russia content – in filling the gap left by banned outlets.

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, in turn, allows us to explore the consequences of reducing exposure to media slant, providing a new dimension to the policy debate on media regulation. Particularly, our setting creates an opportunity to study adjustments in the media market with respect to both the supply of slanted content and its demand, measured as users' engagement with propaganda 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 establishing that the removal of suppliers of slanted (pro-Russia) content can also reduce production among other suppliers by increasing the production cost of content. These remaining suppliers are not able to compensate for the overall reduction in supply of slant and also do not experience an increase in engagement in the very short run.

The remainder of the paper is structured as follows. Section 2 provides background on the ban and presents our conceptual framework. Section 3 discusses the data and the empirical methods. We present our results using observational data in Section 4 and refer to additional results in Section 5. We describe our experimental evidence on the cost of censorship in Section 6 and Section 7 concludes the paper.

2 Setting: The Ban of Russia Today and Sputnik

2.1 The Onset of the War

On February 24$^{\text{th}}$ 2022, the Russian Federation invaded 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. 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.[5] From the outset, the war was not only fought on the battlefield but also in the digital sphere. Russian state-backed outlets provided narratives and selective framing aimed at justifying the invasion and undermining support for Kyiv in the European media space.

The European Union took swift action with the goal of countering this strategy and regulating the dissemination of such narratives. On March 1st 2022, the European Counicl's Decision[6] provided the legal basis to suspend the broadcasting activities of Russia Today (all language services) and Sputnik throughout the EU. One day later, a dedicated Council press release confirmed the immediate ban across television, radio, satellite, cable, internet service providers, and social-media platforms. The Council justified the measure on the grounds that both outlets were "essential and instrumental" in supporting Moscow's aggression and constituted a direct threat to public order and security within the EU.

In this section, we address two questions that set the stage for our analysis. First, why did the EU single out these two outlets? Were they truly influential in shaping wartime discourse, particularly on Twitter? Second, was the ban effectively implemented and were there measurable effects on users' ability to access and engage with the content produced by Russia Today and Sputnik?

2.2 Motivation and Efficacy of the Ban

> *"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*

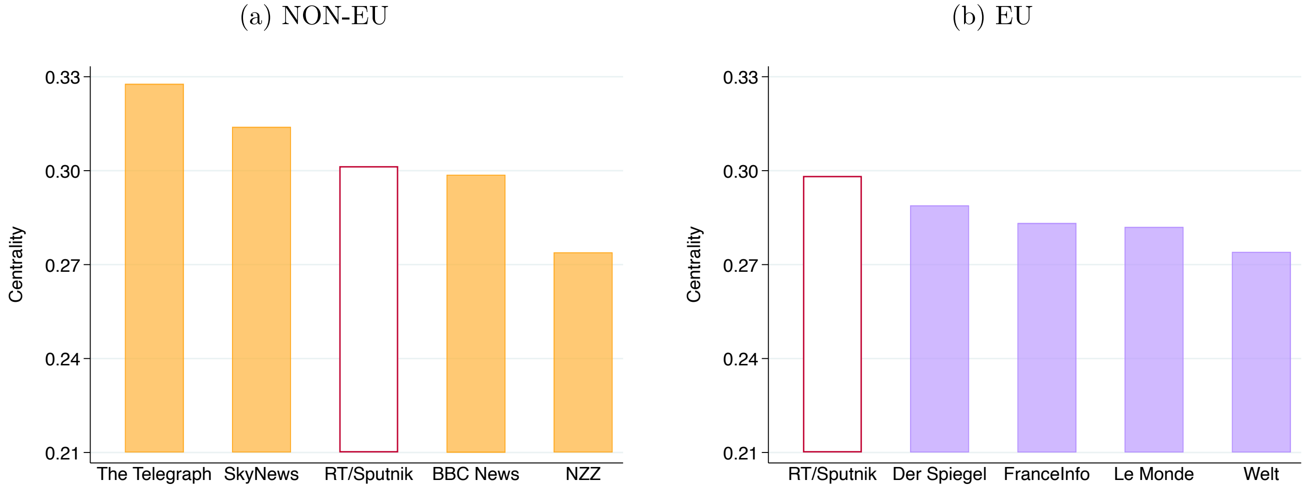
This statement by Josep Borrell – then High Representative of the Union for Foreign Affairs and Security Policy – conveys the EU's assessment of Russia's information campaign at the time of the invasion. It frames propaganda and disinformation as operational weapons in the assault on Ukraine and as direct threats to the Union's security. The proposed logic is straightforward: safeguarding the EU's information space requires strong action. Closing the communication channels of Russian state-controlled media was seen as a key step to counter online warfare, with Russia Today and Sputnik considered the most prominent of these channels. Below, we ask whether these outlets were in fact relevant players by comparing the centrality of Russia Today and Sputnik within the conversation about the war to some selected EU and non-EU news outlets.

Figure I compares the prominence of the banned outlets with that of other major news sources. The underlying dataset is the same used throughout our analysis and contains only tweets about the war, identified through the keyword-filtering procedure described in Section 3. To construct the figure, we extract every user mention (strings beginning with '@') 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 quickly information from that account can reach the rest of the network. Accounts with more

[5] See the article by George C. Marshall: European Center for Security Studies.
[6] Decision (CFSP) 2022/351 and Regulation (EU) 2022/350.

FIGURE I

CENTRALITY OF RUSSIA TODAY AND SPUTNIK VS. SELECTED NEWS OUTLETS



Notes: The figure compares the centrality of Russia Today and Sputnik, treated as a single entity, to the centrality of other selected news outlets within our areas of analysis. We use the so-called closeness centrality, calculated as follows: we compute the network formed by all the tweets used in the analysis by gathering all tweets that mention any user, i.e., when a user is explicitly referred to using the symbol "@" by another user. Counting mentions in this way includes retweets and replies as well as any other way of "@-mentioning" a user. In the next step, we exclude self-mentions to form a directional network with users as nodes and mentions as edges. The closeness value C of a user u is the reciprocal of the shortest path distance d to u over all other connected nodes v: $C(u) = [1 - n]/[\sum_{v=1}^{n-1} d(v, u)]$. Intuitively, a user with a larger closeness value can spread information more quickly through the network than a user with a smaller closeness value. A user with a closeness value of 1 would be able to directly reach every other user in our sample. On the flip side, a user with a closeness value of 0 would be completely unconnected to any other user in our sample. Since the tweets all concern the war, closeness captures the centrality of each user within the war-related discussion. We compute two separate closeness measures: one using tweets originating from the non-EU countries in our study (the UK and Switzerland), and another using tweets from the EU countries in our sample (Austria, France, Germany, Ireland, and Italy). Panel (a) compares the closeness centrality of Russia Today and Sputnik, in the red frame, to that of other non-European news outlets, in orange, using the closeness centrality measure derived from EU-country posts. Panel (b) presents a similar comparison between Russia Today and Sputnik and selected news outlets, in violet, from the EU countries. The panels use the same scale on the y-axis to facilitate comparison.

mentions, and thus a more central role in the conversation, score higher on this measure. The figure compares the centrality of Russia Today and Sputnik (pooled as a single entity) with that of leading news outlets in non-EU countries in our sample (the United Kingdom and Switzerland, left panel) and in EU countries (Austria, France, Germany, Ireland, and Italy, right panel).

The figure shows that, before the ban, Russia Today and Sputnik held a prominent position in war-related conversations in both non-EU and EU countries. In non-EU countries, their centrality was only slightly lower than that of major outlets such as *The Telegraph* and *Sky News*, and even higher than *BBC News*. In EU countries, they ranked above well-established sources like *Der Spiegel* and *Franceinfo*. These descriptive patterns support the EU's assessment: before the ban, Russia Today and Sputnik were significant actors in shaping Twitter discourse on the war.

Given the evidence of the outlets' prominence in the war conversation, we next ask whether the ban effectively reduced the activity of Russia Today and Sputnik in the EU. Figure II provides an initial "first-stage" test. Using the same corpus of war-related tweets described in Section 3 and underlying

Figure I, we count how often users mentioned the two outlets before and after the ban. Counts are shown separately for non-EU countries in our sample (orange) and EU countries (purple). While mentions are only an indirect proxy for accessibility,[7] they remain the best available indicator given Twitter's data constraints.

The figure shows a stark contrast between non-EU and EU countries. In non-EU countries, mentions of the banned outlets surged after the ban, nearly doubling in number. This reflects the broader media dynamic at the time: the ban came about a week after the full-scale invasion, when discussion of the war – and the activity of both outlets – was already intensifying. In the EU, the pattern is the opposite: mentions dropped by more than half once the ban took effect. Importantly, the EU measures blocked access to the outlets' content but did not remove their Twitter accounts. Users could still tag @RT or @Sputnik, but could not view or share their material. In effect, residual mentions from within the EU primarily signal an ongoing willingness to refer to the outlets despite not having access to their content.

Here, a potential concern is that the residual mentions are a signal that the ban was not properly enforced or that there were ways to circumvent the ban. As outlined in detail in Appendix A.2, we use Twitter users' locations as indicated in their profiles to assign them to a country. These locations are public information and can be retrieved via the Twitter API. However, Twitter has also been gathering non-public information on users' locations to determine the content that cannot be displayed to a user, such as in the case of the EU's ban. According to Twitter's public documentation, such information does not only include IP addresses, which can be easily changed via a VPN, for example, but also wireless networks or cell towers near a user. Crucially, manually changing this non-public country setting does not change the content withheld by Twitter due to local laws. Therefore, even if users who interacted with RT and Sputnik who reside in one of the countries that enacted the ban and had their location assigned by Twitter use a VPN to reach the website of the outlets, they still cannot access, read, or interact with any account of RT and Sputnik.

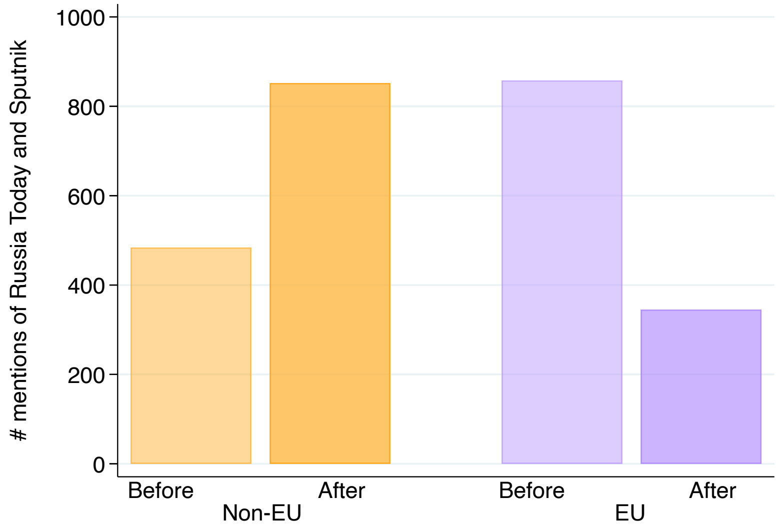
Overall, this descriptive evidence paints a clear picture. Before the ban, Russia Today and Sputnik were central to Twitter discussions of the war. The ban then sharply reduced their online visibility: mentions of the two outlets in EU tweets fell by more than half. Although we cannot do a systematic assessment of the ban's enforcement, our proxy, the tweet-level mentions, indicates that implementation was largely successful. However, estimates should be interpreted as intent to treat estimates as we cannot rule out that circumvention was completely impossible.

2.3 Conceptual Framework

This subsection lays out a minimal conceptual framework to structure the empirical analysis and interpret the results. Our focus is on the behavior of standard users – non-institutional Twitter accounts located inside or outside the EU – who generate war-related tweets. The guiding question is twofold: (i) how did Russia Today (RT) and Sputnik shape the production of pro-Russia content before the ban, and (ii) through which channels might their sudden removal affect that content afterwards?

[7] "@" handles can be used to mention an account even if the account itself is not accessible.

FIGURE II
MENTIONS (@) OF RUSSIA TODAY AND SPUTNIK BEFORE AND AFTER THE BAN



Notes: The figure shows how frequently the handles ("@") of Russia Today and Sputnik, combined into a single entity, appear in tweets from the non-EU (the UK and Switzerland) and EU countries (Austria, France, Germany, Ireland, and Italy) used in our analysis. Frequencies are presented as absolute numbers separated by origin and time relative to the ban. Note that even after the ban, users located in the EU could still mention the outlets' handles in their tweets; however, they could no longer access or retweet content from these accounts.

Benchmark for the effectiveness of the ban EU officials stated that the policy goal was to "turn off the tap" of Kremlin-backed disinformation. Thus, in our setting, a successful implementation of the ban should translate into a measurable decline in the amount of pro-Russia content produced by EU-based users. How, exactly, such a decline might (or might not) arise is ultimately an empirical question. Nevertheless, below we provide some potential mechanisms and scenarios through which the ban could have impacted the media market.

Outlets as reference suppliers We argue that before the ban, RT and Sputnik functioned as reference suppliers of Kremlin-aligned narratives with three complementary mechanisms central to understanding their role before the ban:

(i) **Agenda-setting**. By posting quickly and in high volume, the outlets steered attention, introduced talking points, and defined which aspects of the war users deemed newsworthy.

(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.

(iii) **Search-cost reduction**. The outlets delivered a steady stream of ready-made text, images, and videos that could be paraphrased or translated, thereby lowering the effort required for ordinary users to produce pro-Russia tweets.

Despite not being comprehensive, these mechanisms provide useful guidance to think about the function the banned outlets fulfilled. Taken together, the removal of the outlets then not only eliminated two prolific voices on Twitter, but fundamentally disrupted the network of information flow. This disruption resembles withdrawing a major supplier from a conventional market: the cost of obtaining a particular "good" (pro-Russian narrative frames) rises, and the structure of supply and demand will adjust in response.

Expected effect of the ban Ex ante, the effect of the ban is unclear. We argue there are three core reasons why one would expect the ban to be effective in reducing pro-Russian content on Twitter. First, the ban eliminates key suppliers, inducing higher search/production costs for pro-Russia narratives. This first-round effect can lead to a decrease in pro-Russian content on Twitter. As users have to exert more effort to find and produce pro-Russian narratives, overall pro-Russia content declines. We expect this to be particularly pronounced among users who relied on the outlets as a source directly, i.e., were directly connected to the outlets before the ban, whereas the effect might be more limited among users who are only indirectly connected to the outlets. Second, the ban might signal stricter moderation of pro-Russian content in the future and thereby increase self-censorship among users. Third, the loss of "seal of approval" might deprive users of a clear editorial line to follow. Greater uncertainty and effort to align with Kremlin messaging could reduce pro-Russia content. Expressed differently, the ban changed the production function of pro-Russian content in the EU.

On the other hand, there are also potentially offsetting reactions by users, other suppliers of pro-Russian content, or the Russian government itself. First, users might denounce or actively resist the ban. Posting pro-Russia content becomes an act of protest against censorship. Second, the market for pro-Russian content might adjust. Secondary suppliers (smaller pro-Russia accounts, foreign outlets, or fringe media) exploit the vacant space of the banned outlets by shifting their slant or increasing activity. Users might shift to these new sources and keep or increase their own production of pro-Russian content. Crucially, this would increase the demand for secondary suppliers and be picked up by increased engagement for pro-Russian content produced by other outlets after the ban is implemented. Third, there might be institutional adaptation by the Russian government itself. Other state-linked actors (e.g. diplomats, cultural institutes, bot networks) or non-banned government media outlets might expand their activity and receive more attention by users – increasing the overall pro-Russian content and engagement with it.

These channels are not mutually exclusive; several may operate simultaneously. Which effect prevails is ultimately an empirical question. In the following, we test which channels dominate by measuring changes in both the supply (tweets produced) and the demand (engagement) for pro-Russia content after the ban.

3 Data and Empirical Method

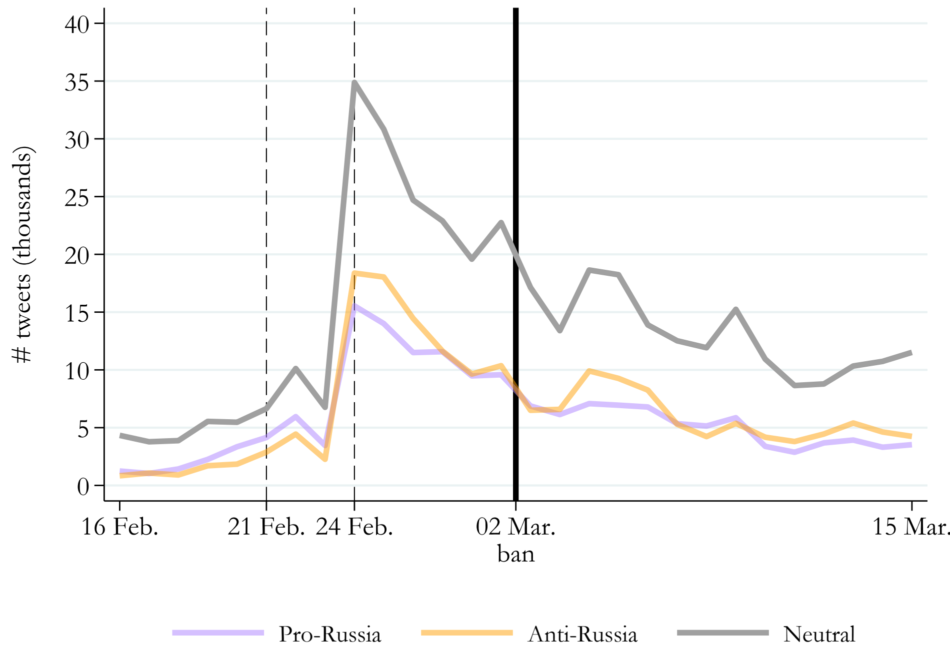
3.1 Measuring Pro-Russia Content

To assess the effectiveness of the ban, we first ask: How can we define and measure "pro-Russia" content on social media? Given the nuance and complexity of the classification task, traditional Natural Language Processing (NLP) methods – such as dictionary approaches, topic modeling, or language similarity – are too limited and risk producing noisy results. To address this, we operationalize the task using recent advances in AI, classifying tweet content with a large language model (LLM). Specifically, we use OpenAI's GPT-4o-mini, accessed via the OpenAI API, to obtain model-generated classifications for our dataset. A key step in this process is the design of the prompt, which has to find the right balance: it should give the model enough freedom to interpret the content, while clearly defining the context and specifying what "pro-Russia" means in our setting. We refine this prompt through iterative testing and adjustments using ChatGPT. Below, we report the first part of the final prompt:

> *"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 strong zero-shot and few-shot performance of GPT-3 and similar models (Brown et al. 2020) has motivated a growing literature on optimal prompt design (Liu et al. 2023). In particular, prior work shows that the set of possible output labels can affect both the predictions and their stability (Zhao et al. 2021). Building on this insight, we instruct the model to classify each tweet into one of three mutually exclusive categories: pro-Russia, neutral, or anti-Russia, specifically in the context of the war. For our main analysis, we focus solely on whether content is pro-Russia or not, using the other categories exclusively for some robustness checks. Figure III reports the distribution of all three labels in the dataset used for our analysis. We provide further details on the tweets dataset in Section 3.2. We also validate this discrete measure via an alternative, continuous measure of pro-Russia slant. We discuss the validation briefly in Section 5 and extensively in Appendix F. Finally, as part of the classification, we also ask the model to list up to five topics mentioned in each tweet. We do so, to leverage the complexity of natural language and to test some of the hypotheses in our conceptual framework.

FIGURE III
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 16$^{\text{th}}$, 2022, to March 15$^{\text{th}}$, 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). From this point onward, the analysis includes only tweets from February 22$^{\text{nd}}$, 2022, to ensure broad coverage and sufficient volume for statistical power in regression models. Vertical lines highlight key dates: February 21$^{\text{st}}$, marking Putin's official recognition of the Donetsk People's Republic and the Luhansk People's Republic; February 24$^{\text{th}}$, indicating the beginning of the war; and March 2$^{\text{nd}}$, representing the start date of the ban.

3.2 Sample of Tweets

Our main analysis sample comprises 677,780 tweets from 146,633 users, posted between February 22$^{\text{nd}}$, 2022, and March 15$^{\text{th}}$, 2022. Due to limitations in accessing the Twitter API,[8] we restrict our extraction to the following European countries: Austria, France, Germany, Ireland, Italy, Switzerland, and the United Kingdom. We access all type of tweets, either original 'own' tweets or replies, or retweets with no more than 140 characters. We discard from our sample retweets of longer original tweets as they are truncated at the 140 character cutoff when retrieved via the API to avoid systematic measurement error in our natural language processing pipeline. We use the sum of retweets, replies, and likes to a given original tweet as our measure for engagement and indication of demand. Additionally, we collect all 'own' tweets posted by Russia Today, Sputnik, and subsidiaries during the same time period.

We show descriptive statistics on the tweet level in Appendix Table B.2, with Panel A focusing on the tweets by Russia Today and Sputnik, and Panel B showcasing our main sample from common users.

[8] Free API access for researchers was unfortunately suspended and the retrievable data changed shortly after Elon Musk's takeover of Twitter in October 2022, during this research project.

In Panel A, we show that Russia Today was more active on Twitter than Sputnik in the time-frame of interest.[9] Additionally, we highlight that news about the Russo-Ukrainian conflict makes up a substantial share of tweets by the outlets, with around 20% of the tweets covering the conflict with a pro-Russian slant. The outlets mention on average 2.2 topics in their tweets.

In Panel B, we show that among all tweets about the Russo-Ukrainian conflict by standard users, the share of pro-Russian content is comparable to the share we find in the outlets' tweets. Similarly, the number of topics about the conflict is also comparable across both panels. Of all users, 15% ever had a connection to Russia Today or Sputnik in the form of a retweet or reply before the ban. Finally, we show that the majority of the users in our sample are located in the UK, France, Germany, and Italy. Overall, the ratio of treated to untreated users is roughly 2:1.

3.3 Identification Strategy

The context of the ban offers a quasi-experimental setting, allowing us to treat it as a natural experiment. Users in the European Union did not anticipate such a swift institutional response, and the ban arrived as a shock that immediately disrupted the European media sphere.[10] Comparable measures were adopted by the United Kingdom and several other countries roughly three weeks later. This timing makes the two-week period immediately following the EU ban a unique window for identifying the short-term effects of censorship in a democratic context. Our identification strategy builds on this setting.

We use a difference-in-difference design to estimate the causal effect of banning the Russian outlets by comparing tweets posted by users in the EU, who were affected by the ban, to tweets posted by users outside the EU, whose exposure to these outlets was not restricted. We estimate difference-in-difference specifications at the user-day level of the following form:

$$Y_{i,t} = \alpha_i + \gamma_t + \beta \; EU_i \times Ban_t + \epsilon_{i,t}, \tag{1}$$

where $Y_{i,t}$ is a measure of user behavior, such as the total amount of pro-Russia tweets produced on day t or a binary variable capturing whether any pro-Russia tweet content was posted, by user i on day t. EU_i is an indicator variable equal to 1 for users located in the European Union and 0 otherwise. Ban_t is an indicator equal to 1 including and after March 2$^{\text{nd}}$, 2022. α_i and γ_t are a full set of user and day-fixed effects absorbing average differences in tweet content across users and time. When we estimate difference-in-difference specifications with outcome variables as counts via Poisson-Pseudo Maximum Likelihood (PPML), then Equation 1 takes the following form:

$$\log \mathbb{E}[Y_{it} \mid \alpha_i, \gamma_t, EU_i \times Ban_t] = \alpha_i + \gamma_t + \beta \; EU_i \times Ban_t. \tag{2}$$

[9] Russia Today used at least six different accounts, tweeting in different languages, namely *@RTenfrancais*, *@de_RT_com*, *@RT_com*, *@de_RT_com*, *@RTUKnews*, and *@RT_America*, while Sputnik only used one account, *@SputnikInt*.

[10] 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). Notably, 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). As noted earlier, the European Court of Justice upheld the ban on the grounds of the prohibition of "propaganda for war".

To interpret β in Equations 1 and 2 as the causal effect of the ban on Russia Today and Sputnik, we require the assumption that Twitter user behavior in EU and non-EU countries would have followed parallel trends in the absence of the ban. This assumption is not directly testable, but we provide evidence supporting parallel trends by estimating an event study specification. For our OLS regressions, this alters Equation 1 to:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k=-8, k\neq -1}^{13} \beta_k \ (EU_i \times Ban_{i,t}^k) + \epsilon_{i,t}, \tag{3}$$

while Equation 2 takes the following event study form:

$$\log \mathbb{E}\Big[Y_{it} \mid \alpha_i, \gamma_t, \{EU_i \times Ban_{i,t}^k\}\Big] = \alpha_i + \gamma_t + \sum_{k=-8, k\neq -1}^{13} \beta_k \ (EU_i \times Ban_{i,t}^k). \tag{4}$$

With this exercise, we compare content posted by users in the EU with that of users outside the EU. We explore whether the content followed similar trends before the ban. We find no meaningful differential trends in outcomes before the ban, making it less likely that the trends would have diverged in the absence of the ban.

Another potential concern is that the invasion itself might impact the use of pro-Russian content on Twitter. Day fixed effects absorb any common shock to EU and non-EU countries. We only include major Western European countries in our sample to make it less likely that differential exposure to the war itself or fear of spillover to the country of residence confounds our estimates. Eastern European countries bordering Russia or Ukraine might have a differential response to the war, so we excluded them from the analysis.

A final concern is that Twitter did not properly enforce the ban, and a clean assignment of the treatment status in our setting also depends on the absence of ways to circumvent the ban. As outlined in detail in Appendix A.2, we assign Twitter users to a country and, thereby, a treatment status based on information in their profiles. We retrieve the locations indicated in the profiles via the Twitter API. However, Twitter also gathers non-public information on users' locations to determine the content that a user cannot retrieve. According to Twitter's public documentation, such information does not only include IP addresses, which a user can easily change via a VPN, for example, but also wireless networks or cell towers near a user. Crucially, manually changing this non-public country setting does not affect the content Twitter withholds due to local laws. Therefore, even if readers of RT and Sputnik who reside in one of the countries that enacted the ban and had their location assigned by Twitter within that country use a VPN to reach the website of the outlets, they still cannot access, read, or interact with any account of RT and Sputnik on Twitter.

4 Main Results: The Impact of the Ban on Pro-Russia Content

4.1 Descriptive Evidence

The first step of our analysis consists of a descriptive exploration of the effects of the ban. As we argued in the sections above, the context in which the ban was implemented and the decision-making process behind it provide a valuable quasi-experimental setting. Although the full-scale invasion – launched on the 24$^{\text{th}}$ of February – had already triggered debates on how to counter Russian misinformation and propaganda, the actual decision to ban the two outlets came unexpectedly. It is reasonable to assume that most users did not anticipate the move and had made no immediate plans to replace these sources of pro-Russia content. In this section, we track how pro-Russia content evolved within Europe, and show descriptive comparisons of the patterns before and after the ban in EU versus non-EU countries.

We begin by presenting descriptive evidence on the effect of the ban. The left panel of Figure IVa plots the daily average number of pro-Russia tweets posted by users in the EU countries in our sample. The graph follows a simple event study design: the vertical line marks the date of the ban's implementation, and the plotted coefficients represent daily averages for EU-based users. The right panel, Figure IVb, shows the share of pro-Russia content before and after the ban, separately for users in non-EU countries (orange) and EU countries (purple), with 95% confidence intervals around the group means. Both figures are based on our user-day panel dataset, restricted to user-days with at least one tweet about the war – whether neutral, pro-Russia, or anti-Russia. To capture pre-trends, we include the seven days before the ban – from February 22$^{\text{nd}}$ onward.

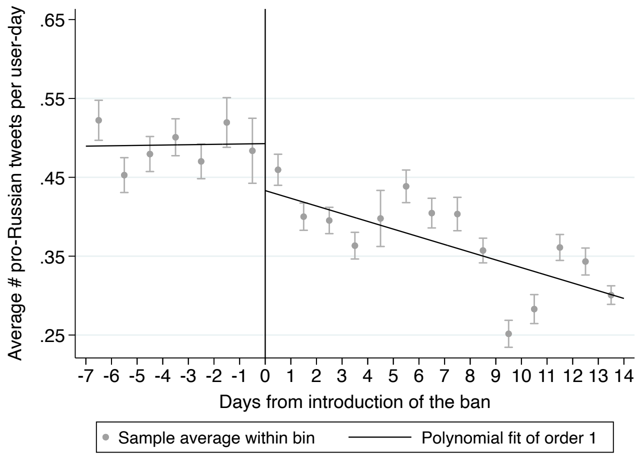
The event study in Figure IVa offers a clear visual account of the ban's effect on users in European countries. In the days leading up to the ban, these users consistently posted between 0.45 and 0.55 pro-Russia tweets per day on average, conditional on being active positing about the war. The trend is broadly stable – if anything, slightly increasing – until the ban is introduced. Immediately after the ban, we see a clear and sudden drop in the average number of pro-Russia tweets per user-day among EU users. Over the following two weeks, the average drops further steadily, reaching about 0.30 pro-Russia tweets per user-day by the end of the period.[11] Figure IVb complements this pattern by plotting the share of pro-Russia content relative to all tweets by active users. While a mild decrease is also visible among non-EU users, the drop in the EU sample is, descriptively, at least three times larger.

Although purely descriptive, these patterns suggest that the ban had a substantial impact on EU-based users. By removing two primary sources of pro-Russia content, the ban likely disrupted users' ability to access material that could serve as inspiration or reinforcement. The resulting decline in pro-Russia output is consistent over time in the two weeks following the ban, with a difference in share of pro-Russia content nearly three times as large as the change observed among non-EU users. In the upcoming sections, we explore the causal effect of the ban using the identification strategy outlined above, investigate heterogeneous effects among subpopulations, potential substitution dynamics, and, finally, shed light on the mechanisms behind the ban's effects.

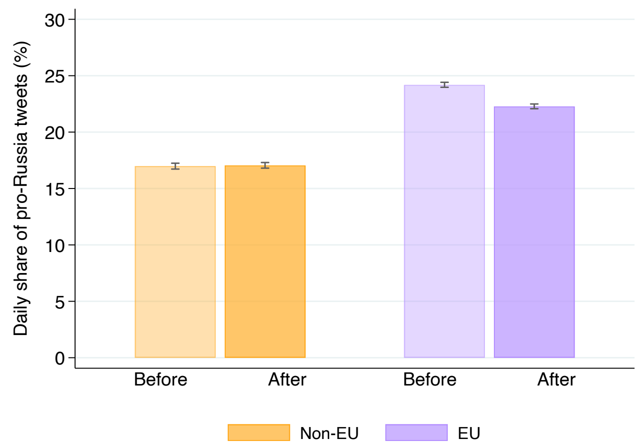
[11] In Appendix Figure C.1 we compare this to the event study for users from non-EU countries. The comparison exercise provides descriptive evidence that there is no jump around the ban in that case.

FIGURE IV
DESCRIPTIVE EVIDENCE OF THE IMPACT OF THE BAN

(a) EVENT STUDY:
PRO-RUSSIA CONTENT IN EU



(b) AVERAGE SHARE
OF PRO-RUSSIA CONTENT



Notes: The figures provide descriptive evidence of the impact of the ban on pro-Russia content in EU countries. Panel (a) presents a regression discontinuity design (RDD) analysis, where the running variable is time. The coefficients trace the evolution of the absolute number of pro-Russia tweets posted by users located in EU countries around the introduction of the ban. The vertical axis shows the daily average, and we present a linear fit of the data on either side of the cutoff. The horizontal axis reports the number of days relative to the ban (day 0), ranging from 7 days before to 14 days after its implementation. Shaded areas denote 95% confidence intervals. Panel (b) displays the share of content classified as pro-Russia in non-EU countries (the United Kingdom and Switzerland) versus EU countries (Austria, France, Germany, Ireland, and Italy), before and after the ban. For both panels, we use the user-day panel dataset and include only user-day observations where the user was active in our sample as described in Section 3.2. We include only observations between February 22nd and March 15th, to ensure a large enough sample size for each day. Appendix Figure C.1 shows the same event study reported in Panel (a), for Non-EU countries.

4.2 General Impact of the Ban

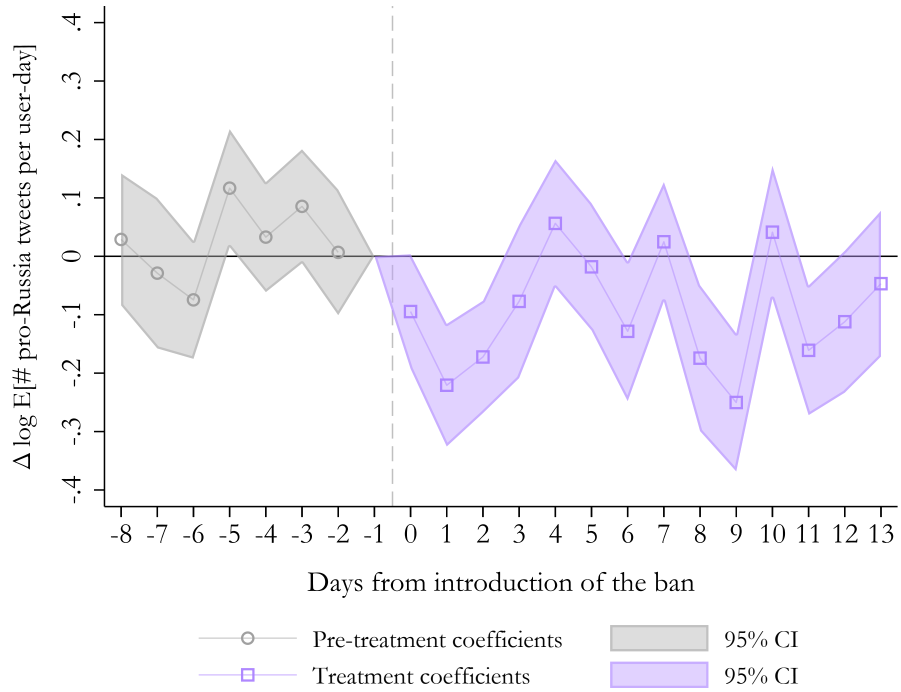
The impact of the ban could have unfolded in the European media market in many ways, and different subpopulations of users may have reacted differently. Nevertheless, the most intuitive starting point is to consider whether it produced a broad, generalized decline in pro-Russia content within the EU, relative to the non-EU market. We begin our regression analysis with this aggregate perspective. At this stage, we do not distinguish among different types of users. The results should thus be interpreted as capturing the overall effect of the ban on the general population of users posting from European countries.

Figure V reports estimates from a daily event study version of Equation 2 using the absolute number of pro-Russia tweets as the dependent variable, and specifying it with a Poisson-Pseudo Maximum Likelihood (PPML) model. For the analysis, we use our user-day panel dataset, including for each user only those days where the user was active in our sample as described in Section 3.2. The analysis includes tweets starting from February 22nd, 2022, up until March 15th, 2022. Each regression incorporates user- and day-fixed effects, with standard errors clustered at the user level.

This event study provides three key insights. First, we do not observe clear pre-trends in the period leading up to the ban. If anything, on two of the days before the ban, we find significantly different

FIGURE V

DAILY EVENT STUDY: IMPACT OF THE BAN ON PRO-RUSSIA CONTENT



Notes: The figure displays coefficients and 95% confidence intervals from estimating Equation 4. The model employs the user-day panel dataset and includes only user-day observations where the user was active in our sample as described in Section 3.2. 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; thus, the coefficients capture the daily average effect of the ban on a user's pro-Russia tweet activity, conditional on the user being active that 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. In Table I, we report the percentage change for the corresponding difference-in-difference estimation by transforming the coefficients β as $e^{\beta} - 1$. The reported regression includes user and day fixed effects, and 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.

coefficients in the opposite direction, indicating a higher volume of pro-Russia content among EU users. Second, immediately after the ban's implementation, the estimates show a negative shift consistent with a decline in pro-Russia content among users in the EU, affected by the ban. Third, although the results are somewhat noisy – reflecting the inherent variability of a user-day panel from a social media platform – the negative effect appears consistent and persistent over time. Taken together, these results suggest a clear impact of the ban on the general population of EU-based users posting about the war.

Using the number of pro-Russia tweets posted in EU versus non-EU countries as the dependent variable has some limitations. First, the estimates may be driven by a small set of users whose behavior changed sharply after the ban. For example, users who previously produced a high volume of pro-Russia content may have been especially responsive to the ban, hence disproportionately influencing the results. Second, absolute counts may hide meaningful changes. Consider two users: one who posted a single pro-Russia tweet before the ban and another who posted fifteen. If both reduce their output by one tweet, the first user has completely stopped producing pro-Russia content, whereas the second has barely changed.

TABLE I
IMPACT OF THE BAN ON USERS' CONTENT

Dependent variable	**# tweets pro-Russia**	**P(Tweet: Pro-Russia)**	**P(Tweet: About war)**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Ban x EU	-0.116 (0.028) [0.000]	-0.027 (0.003) [0.000]	-0.007 (0.001) [0.000]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	
Pre-ban outcome avg. for treated	0.487	0.323	0.161
Approx. percentage change	-10.92	-8.49	-4.46
Observations	215723	322199	3325498

Notes: The table presents the results from two-way fixed effects difference-in-difference regression models analyzing the impact of the ban on users' posting behavior. All models use the user-day panel dataset and include observations from February 22$^{\text{nd}}$ to March 15$^{\text{th}}$, 2022. Column 1 reports estimates on the impact on the total number of pro-Russia tweets via Equation 2, conditional on tweeting about the war. Column 2 shows estimates of posting pro-Russia content conditional on tweeting about the war using Equation 1. Column 3 reports estimates of posting any war-related content, again via Equation 1. We compute the approximate percentage change in Column 1 as $e^{\beta} - 1$, and in Columns 2 and 3 as the change from the pre-ban average. All specifications include user and day fixed effects, and standard errors are clustered at the user level.

Third, the ban may affect not the way users talk about the war, but whether they talk about it at all. In that case, the decline in pro-Russia content would reflect a withdrawal from the topic rather than a shift in stance. We address these issues in the results reported in Table I.

Table I reports estimates from regressions using specification 1. Column 1 uses the total number of pro-Russia tweets as the dependent variable, conditional on posting about the war; in effect, it summarizes in a two-period regression the effects shown in the event study above. Column 2 replaces the dependent variable with an indicator for posting any pro-Russia content, again conditional on posting about the war. Column 3 uses as the dependent variable an indicator for posting about the war at all, hence capturing changes in participation in the discussion irrespective of stance. For all models, the table also reports the pre-ban average of the dependent variable among treated users.

The results offer a clear and consistent picture of the ban's effects. In line with the more granular event study estimates, the ban leads to a marked reduction in pro-Russian content: among EU users, the total volume of such tweets on active user-days decreases by approximately 11% relative to non-EU users. Quantifying this effect in Appendix Table C.1, the ban led to a reduction of over 3000 pro-Russia tweets – own tweets, replies, or retweets – per day among users that we can locate in the EU. To ensure this decline is not driven by a small number of users drastically altering their behavior, Column 2 explores the effect on the probability of posting pro-Russian content, conditional on tweeting about the war on a given day. The decline amounts to roughly 8.5% post-ban, reinforcing the notion that the effect is widespread rather than concentrated among a few individuals. Finally, we ask whether the ban results in users withdrawing from the conversation altogether, rather than changing the tone of their posts. We find that the ban does reduce the likelihood of tweeting about the war on a given user-day, the drop is comparatively modest, with about a 4.5% decrease relative to pre-ban averages.

It is important to emphasize that this analysis speaks to the supply side of online political discourse. Specifically, it asks: Does banning two major suppliers of pro-Russian narratives reduce the presence of such content among ordinary users? As a first-round effect, the ban results in a decreased presence of the outlets in the EU discussion about the war. In addition, the evidence presented in this section suggests that the ban reduces pro-Russian content among ordinary users as a secondary effect. European users, who were subject to the ban, not only produced fewer pro-Russian tweets in absolute terms but also became less likely to post such content when participating in the discussion on active days. In the following section, we disentangle the aggregate effect in more detail and turns to the role of proximity to the banned outlets. If these outlets played a central role in shaping narratives and influencing users, we would expect those more closely connected to them to exhibit stronger behavioral responses to the ban.

4.3 The Impact on the Outlets' Network

The next step of our analysis focuses on users who were – directly or indirectly – part of Russia Today's and Sputnik's network on Twitter, a group we label *connected users*. A user is classified as connected if she (i) retweeted or replied to RT/Sputnik at any point before or during our study period, or (ii) engaged with another user in this network, at any degree of separation. This group matters for several reasons. First, the ban directly cut off their access to a central source of information and inspiration for pro-Russia content. Second, although they represent a relatively small share of the sample, they accounted for a disproportionate share of tweets about the conflict. Finally, as shown in Appendix Figure B.2, they were far more likely to disseminate pro-Russia narratives before the ban.

As a counterpart to connected users, we also examine those who never engaged with Russia Today or Sputnik, either directly or indirectly. Analyzing this group serves two purposes. First, it provides a benchmark against which to interpret the behavioral shifts observed among connected users. Second, it allows us to assess whether the ban produced broader effects, consistent with the EU's stated aim of limiting the spread of pro-Russian narratives in general. While the outlets themselves were the central targets of the policy, its ambition extended to reshaping the wider information environment, beyond the outlets and their immediate network.

Table II reproduces the main results from Table I, but distinguishes between the two subpopulations described above. The top panel reports estimates for connected users – those who were directly or indirectly engaging with Russia Today and Sputnik – while the bottom panel reports results for non-connected users, who never directly interacted with the outlets or any other ordinary user that was part of the outlets' network. In both panels, the table follows the same structure as before: Column 1 presents results for the absolute number of pro-Russia tweets, Column 2 for the likelihood of posting pro-Russia content, and Column 3 for the likelihood of posting at all about the war. As above, the outcomes in Columns 1 and 2 are measured conditional on users posting about the war.

The comparison highlights several noteworthy patterns. First, for the number of pro-Russia tweets conditional on posting about the war, the effect among connected users is roughly double that among non-connected users: a decline of about 11.6% relative to the pre-ban mean for the former, compared to

TABLE II
HETEROGENEOUS IMPACT OF THE BAN BY CONNECTION TO THE OUTLETS

PANEL A: IMPACT ON USERS CONNECTED TO THE OUTLETS

Dependent variable	**# tweets pro-Russia**	**P(Tweet: Pro-Russia)**	**P(Tweet: About war)**
		Coeff./SE/p-value	
	(1)	(2)	(3)
Ban x EU	-0.124	-0.029	-0.018
	(0.033)	(0.004)	(0.001)
	[0.000]	[0.000]	[0.000]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	
Pre-Ban outcome avg. for treated	0.580	0.580	0.247
Approx. percentage change	-11.69	-5.08	-7.37
Observations	148764	202671	1383602

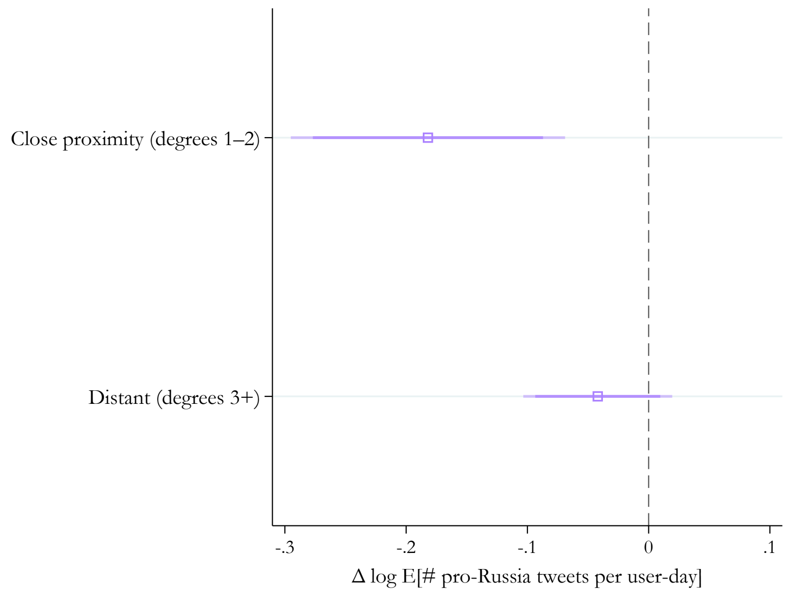
PANEL B: IMPACT ON USERS NOT CONNECTED TO THE OUTLETS

Dependent variable	**# tweets pro-Russia**	**P(Tweet: Pro-Russia)**	**P(Tweet: About war)**
		Coeff./SE/p-value	
	(1)	(2)	(3)
Ban x EU	-0.057	-0.020	0.001
	(0.034)	(0.006)	(0.001)
	[0.099]	[0.000]	[0.342]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	
Pre-Ban outcome avg. for treated	0.323	0.259	0.100
Approx. percentage change	-5.52	-7.69	0.95
Observations	66959	119528	1941896

Notes: The tables report heterogeneous effects of the ban by users' connection to Russia Today and Sputnik. Users are classified as connected if they belong to the outlets' network, meaning they either (i) directly retweeted or replied to RT/Sputnik at any point, or (ii) retweeted/replied to another user who is part of that network, at any degree of separation. Panel A shows results for connected users, Panel B for non-connected users. We present estimates of two-way fixed effects difference-in-differences regressions on user-day data, on the user-day panel dataset from February 22^{nd} to March 15^{th}. For both tables, Column 1 reports estimates on the impact on the total number of pro-Russia tweets via Equation 2, conditional on tweeting about the war. Column 2 shows estimates of posting pro-Russia content conditional on tweeting about the war using Equation 1. Column 3 reports estimates of posting any war-related content, again via Equation 1. We compute the approximate percentage change in Column 1 as $e^{\beta} - 1$, and in Columns 2 and 3 as the change from the pre-ban average. All specifications include user and day fixed effects, and standard errors are clustered at the user level.

about 5.5% for the latter. Moreover, the effect for non-connected users is statistically significant only at the 10% level. A different picture emerges for the likelihood of posting pro-Russia content, conditional on posting about the war. Here, the effect size is similar across groups, with the estimate for non-connected users approximately 2.5 percentage points larger than that for connected users, reaching a value of about 7.7%. Finally, in Panel B, the effect of the ban on non-connected users regarding the likelihood of posting

FIGURE VI
IMPACT ON CONNECTED USERS BY PROXIMITY TO THE OUTLETS



Notes: The figure displays coefficients, 90% confidence intervals (dark violet), and 95% confidence intervals (light violet) from estimating Equation 4 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 use the user-day panel dataset and include only user-day observations where the user was active in our sample as described in Section 3.2. 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 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. 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 E.1, we report, among other details, the number of observations as well as the percentage change for the corresponding difference-in-difference estimation by transforming the coefficients β as $e^{\beta} - 1$. All reported regressions include user and day fixed effects, and standard errors are clustered at the user level.

about the war at all is essentially zero. In Panel A, for connected users, the effect is sizable, a decline of roughly 7.3%.

Overall, the results suggest that the ban's impact extended well beyond the relatively small network of connected users, despite with different intensity of effects. Among non-connected users, the decline in the absolute number of pro-Russia tweets is statistically significant only at the 10% level. However, its magnitude remains substantial – roughly half the size of the effect observed among connected users. The larger impact on connected users is consistent with their higher baseline production of pro-Russia content before the ban. Importantly, when focusing on the likelihood of posting pro-Russia content conditional on tweeting, we find a sizable decline even among non-connected users. This pattern indicates that the ban did have a broader effect extending beyond users who were directly affected by the removal of outlets they engage with.

In the final part of this section, we step further into the network dynamics surrounding the banned outlets. We conceptualize these outlets as major suppliers – central nodes in a network of users who drew inspiration from them and, in turn, spread similar content. If this view is correct, we would expect users closest to the outlets to be most strongly affected by the ban. Figure VI investigates this hypothesis by focusing exclusively on the group of connected users, and subdividing them into two groups depending on their proximity to the central nodes of the network, the two outlets. The figure represents a coefficient plot showing the impact of the ban on the number of pro-Russia tweets, differentiated by users' proximity to the outlets. The first (top) coefficient captures the effect on users with high proximity – defined as those who directly retweeted or replied to content from Russia Today or Sputnik (first-degree connections), or who engaged with users holding such connections (second-degree). The second (bottom) coefficient reflects the effect on users with weaker ties, defined as third-degree connections or further away.

The results are consistent with this network-based view: the ban's impact declines with distance from the targeted outlets. Users most closely connected to Russia Today and Sputnik exhibit the largest reduction in the volume of pro-Russia tweets, while the effect diminishes among more distant users and is negligible for those with no identifiable link. This gradient reinforces the interpretation of these outlets as central hubs for pro-Russia narratives – actors that not only supplied content but also set the agenda and provided stylistic cues for how such narratives should be framed and disseminated.

4.4 Impact on Secondary Suppliers

Up to this point, we have shown that the ban substantially reduced the production and dissemination of pro-Russia content in the EU media market. After its implementation, EU-based users produced and shared fewer pro-Russia tweets in absolute terms and were less likely to post such content at all per active day. As expected, the reduction is largest among users embedded in the outlets' network, yet the impact extends well beyond this group. To better capture this broader effect, we now turn to a more transversal set of users, which we call the *secondary suppliers*.

By secondary suppliers, we mean users who actively produced pro-Russia content during the eight days leading to the ban. A secondary supplier is any user who created at least one tweet – own tweet or reply, excluding retweets – labeled as pro-Russian content before the ban. This group may include, on the one hand, connected users who did not simply consume or spread the outlets' messages but generated pro-Russia tweets themselves, and on the other, unconnected users who produced such content independently. Studying them is crucial: unlike more passive consumers, these users 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.

As a first step, we reproduce our main specifications in Table III, restricting the sample to secondary suppliers. The models resemble what we show in Table I, but focus only on tweets created by the users themselves, excluding retweets. Overall, the ban did not reduce the production of pro-Russia content among secondary suppliers. These users were active producers before the ban and continued to be, even after the outlets' removal. Similarly, we find a null effect on the likelihood of posting pro-Russia content, conditional on being active. In this aggregate view of the suppliers, we see no impact of the ban on the production of pro-Russia content. Finally, in Column 3, we show that the likelihood of suppliers tweeting about the war in general increases by roughly 0.5%. While the size of this effect is negligible, it suggests

TABLE III
IMPACT OF THE BAN ON SECONDARY SUPPLIERS' CONTENT PRODUCTION

Dependent variable	**# tweets pro-Russia**	**P(Tweet: Pro-Russia)**	**P(Tweet: About war)**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Ban x EU	-0.046	0.003	0.005
	(0.042)	(0.002)	(0.001)
	[0.272]	[0.135]	[0.000]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	
Pre-ban outcome avg. for treated	0.344	0.206	0.957
Approx. percentage change	-4.50	1.36	0.55
Observations	100877	522255	538032

Notes: The table presents the results from two-way fixed effects difference-in-difference regression models analyzing the impact of the ban on secondary suppliers' posting behavior. Users are defined as secondary suppliers if they posted at least one pro-Russia tweet in the eight days before the ban. All models use the user-day panel dataset, excluding retweets, and include observations from February 22$^{\text{nd}}$ to March 15$^{\text{th}}$, 2022. Column 1 reports estimates on the impact on the total number of pro-Russia tweets via Equation 2, conditional on tweeting about the war. Column 2 shows estimates of posting pro-Russia content conditional on tweeting about the war using Equation 1. Column 3 reports estimates of posting any war-related content, again via Equation 1. We compute the approximate percentage change in Column 1 as $e^{\beta} - 1$, and in Columns 2 and 3 as the change from the pre-ban average. All specifications include user and day fixed effects, and standard errors are clustered at the user level.

that the ban made some users slightly more active in the war-related discussions on Twitter.

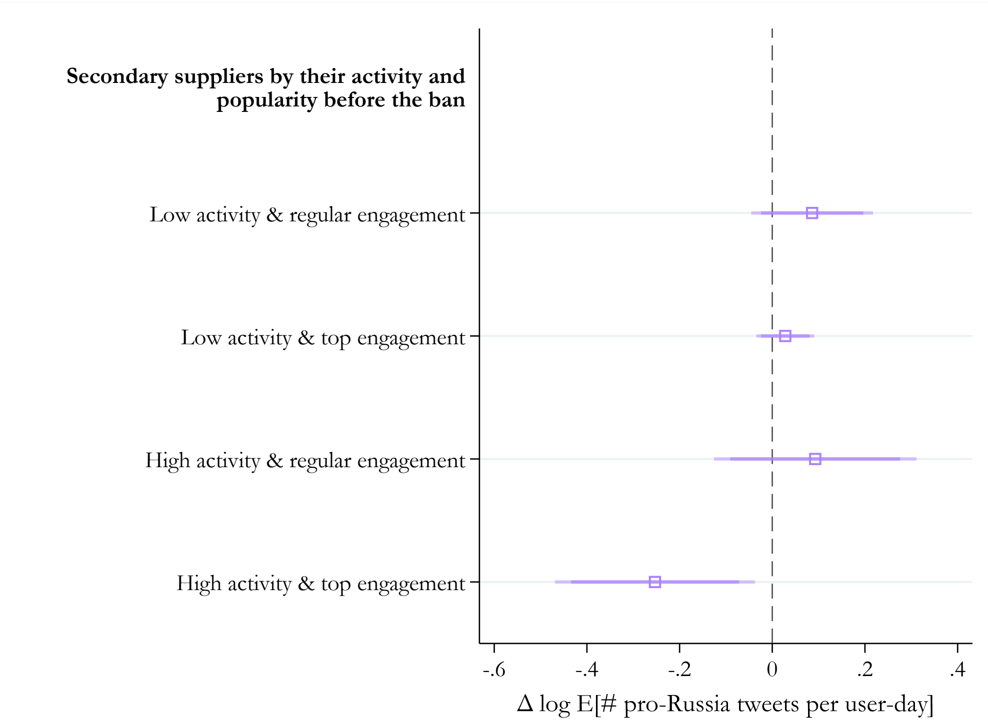
Next, we explore two dimensions of heterogeneity among secondary suppliers: their level of pre-ban activity and the average engagement their pro-Russia content received, before the ban. For activity, we distinguish between low-activity users (three days or fewer of pro-Russia activity over the eight-day window before the ban) and high-activity users (four days or more). For engagement, we define top engagement users as those whose pro-Russia content ranked in the top 10% of the pre-ban distribution of retweets, replies, and likes per tweet.

We tackle this issue in Figure VII, which shows the impact of the ban on the number of pro-Russia tweets produced by the subgroups described above. The analysis captures heterogeneity along the two key dimensions mentioned above: (i) the level of pre-ban pro-Russia activity and (ii) the engagement such content received before the ban. The results paint a clear picture. Most users were unaffected, while others showed declines once Russia Today and Sputnik disappeared. The first two coefficients from the top correspond to suppliers with low pre-ban activity, separated into those with regular and high engagement. For these groups, the ban did not trigger a negative effect. We find a similar effect for highly active suppliers who received regular engagement on their pre-ban pro-Russia content. By contrast, we observe a decline among highly active users whose content had attracted substantial engagement before the ban.

Finally, building on these heterogeneous effects, in Appendix B we provide additional descriptive evidence, helpful to further characterize these subpopulations. Figure B.3a shows that highly active users were disproportionately more likely to be connected to Russia Today or Sputnik, consistent with

FIGURE VII

IDENTIFYING THE SECONDARY SUPPLIERS MOST AFFECTED BY THE BAN



Notes: The figure displays coefficients, 90% confidence intervals (dark violet), and 95% confidence intervals (light violet) estimating Equation 4 while dissecting the heterogeneous effect of the ban on the content production of secondary suppliers. Users are defined as secondary suppliers if they posted at least one pro-Russia tweet (own tweet or reply, excluding retweets) in the eight days before the ban. All regressions use the user-day panel dataset, excluding retweets, and include only user-day observations where the user posted at least one tweet (own tweet or reply, excluding retweets) 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 figure divides secondary suppliers by their activity before the ban and the amount of engagement their pro-Russia content (own tweet or reply, excluding retweets) received before the ban. For activity, we separate suppliers into those who produced pro-Russia content (own tweet or reply, excluding retweets) in three or fewer days before the ban (low activity) and those who produced pro-Russia content (own tweet or reply, excluding retweets) in more than three days before the ban (high activity). With respect to engagement, we distinguish between regular and top engagement. The top engagement includes users whose pro-Russia content received the top 10% number of retweets, likes, and replies per tweet (own tweet or reply, excluding retweets). Regular engagement captures anything else. The dependent variable is the number of pro-Russia tweets (own tweet or reply, excluding retweets) posted per user per day; coefficients therefore capture the daily average effect of the ban on original 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 E.3, we report, among other details, the number of observations as well as the percentage change for the corresponding difference-in-difference estimation by transforming the coefficients β as $e^{\beta} - 1$. All reported regressions include user and day fixed effects, and standard errors are clustered at the user level.

a stronger reliance on the outlets for inspiration. Appendix Figure B.3b similarly shows that greater activity is associated with shorter network distance to the outlets, while Appendix Figure B.4 shows that the most active users received disproportionately higher levels of per-tweet engagement on their pro-Russia content. Finally, Appendix Figure C.3 confirms that the negative impact of the ban intensifies with activity, producing the steepest decline among the most active subgroup and no effect for those who had posted pro-Russia content on fewer days. Taken together, these patterns provide a clearer picture of the users most affected by the ban: they were not casual participants, but rather highly active, well-connected, and more influential accounts at the core of the pro-Russia content network.

4.5 Suggestive Evidence on the Demand Side

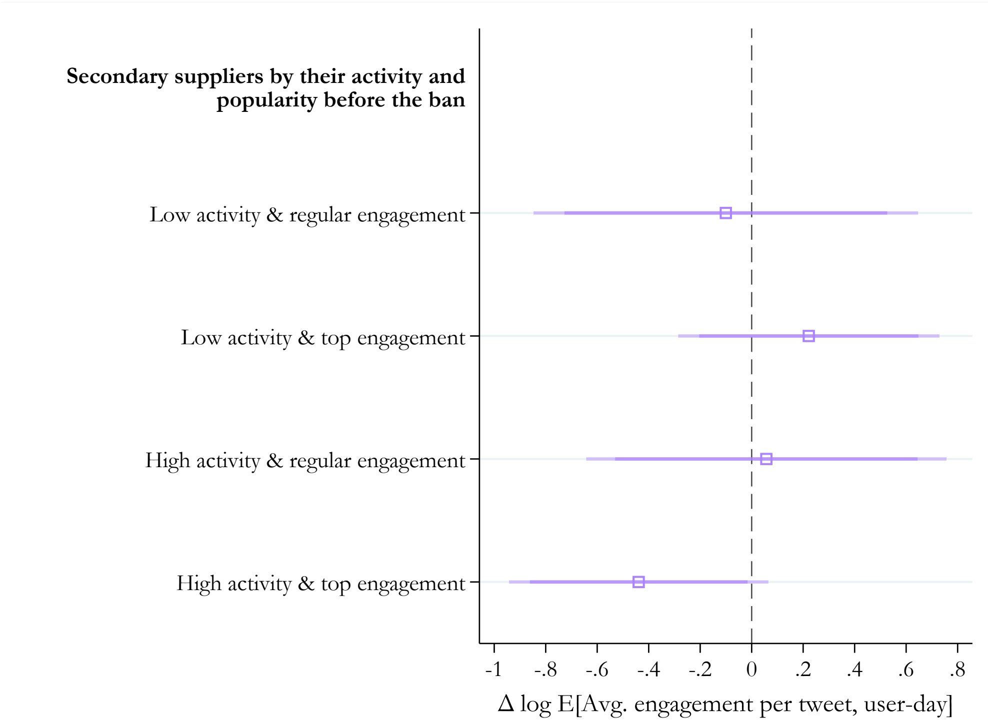
The approach we have adopted so far leaves the demand side of pro-Russia content unexplored, which may be a limitation in our context. The short time horizon of our analysis and the sudden implementation of the ban make it plausible that demand for such content was relatively inelastic. Prior research shows that when fewer actors disseminate a signal, the remaining ones may attract disproportionately more attention and engagement (Germano, Gómez, and Le Mens 2019). Applied to our setting, with fewer users producing pro-Russia content and less content overall, those who remained active may have received greater engagement after the ban than before. At the same time, the ban could also have reduced demand: users searching for pro-Russia content might have left the platform, or the ban could have been interpreted as a signal of social disapproval, discouraging engagement with such messages. In this section, we therefore ask: Do users who continued posting pro-Russia content receive more engagement after the ban?

Answering this question is challenging, particularly given the limitations of our data. While we cannot observe the number of views a tweet receives, we can proxy demand through engagement – measured as the sum of likes, retweets, and replies. Since our interest lies in engagement with pro-Russia content, we focus again on users we label as secondary suppliers – those who produced at least one pro-Russia tweet (own tweet or reply) in the days before the ban. Within this group, our goal is to identify whether certain users experienced an increase in engagement with their pro-Russia content after the ban. Intuitively, users who already attracted high engagement prior to the ban may have seen their visibility and reach expand further once Russia Today and Sputnik were removed. Two mechanisms could explain this. First, users who had previously engaged with the banned outlets may have redirected their attention, deliberately amplifying the content of secondary suppliers once the outlets disappeared. Second, platform dynamics may have played a role: Twitter's algorithm, designed to maximize user engagement, may have boosted 'second-best' alternatives to compensate for the sudden disappearance of prominent pro-Russia accounts, elevating the visibility of users who had already demonstrated traction before the ban.

Figure VIII examines the effect of the ban on the engagement that secondary suppliers' pro-Russia content receives. To connect this demand-side analysis with the earlier supply-side results, we use the same classification of secondary suppliers as in Figure VII. Specifically, we divide suppliers into four mutually exclusive groups based on their level of pro-Russia activity before the ban and the engagement their pro-Russia content received pre-ban. As before, top engagement refers to the top 10% of users by engagement per tweet prior to the ban. We adopt this approach not only for consistency but also because pre-ban activity and engagement are two key dimensions for understanding how the media market adjusted in response to the ban.

The figure reports results from Poisson-Pseudo Maximum Likelihood regressions, where the dependent variable is the total engagement – the sum of retweets, replies, and likes – secondary suppliers received on their average pro-Russia tweets per user-day. Overall, the heterogeneous pattern we find here is similar to that of pro-Russia content in the previous section. Among secondary suppliers with low

FIGURE VIII
ENGAGEMENT OF SECONDARY SUPPLIERS' PRO-RUSSIA CONTENT



Notes: The figure displays coefficients, 90% confidence intervals (dark violet), and 95% confidence intervals (light violet) estimating Equation 4 while dissecting the heterogeneous effect of the ban on the engagement per tweet, which the pro-Russia content of different groups of secondary suppliers receives. Users are defined as secondary suppliers if they posted at least one pro-Russia tweet (own tweet or reply, excluding retweets) in the eight days before the ban. All regressions use the user-day panel dataset, excluding retweets, and include only user-day observations where the user posted at least one tweet (own tweet or reply, excluding retweets) about the war. We include only observations between February 22^{nd} and March 15^{th}, to ensure a large enough sample size for each day. The figure divides secondary suppliers by their activity before the ban and the amount of engagement their pro-Russia content (own tweet or reply, excluding retweets) received before the ban. For activity, we separate suppliers into those who produced pro-Russia content (own tweet or reply, excluding retweets) in three or fewer days before the ban (low activity) and those who produced pro-Russia content (own tweet or reply, excluding retweets) in more than three days before the ban (high activity). With respect to engagement, we distinguish between regular and top engagement. The top engagement includes users whose content (own tweet or reply, excluding retweets) received the top 10% number of retweets, likes, and replies. Regular engagement captures anything else. The dependent variable is the per-tweet engagement – the average number of likes, retweets, and replies – secondary suppliers achieve on their pro-Russia tweets (own tweet or reply, excluding retweets) per user per day; coefficients therefore capture the daily average effect of the ban on per-tweet engagement, 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 engagement. In Appendix Table E.4, we report, among other details, the number of observations as well as the percentage change for the corresponding difference-in-difference estimation by transforming the coefficients β as $e^{\beta} - 1$. All reported regressions include user and day fixed effects, and standard errors are clustered at the user level.

pre-ban activity, as well as for secondary suppliers with high pre-ban activity and regular engagement, we find no change in engagement due to the ban. For highly active users with top pre-ban engagement, we find suggestive evidence for a drop in their pre-tweet engagement on pro-Russia content after the ban. The effect is statistically significant at the 10% level and implies a 34% engagement decrease per pro-Russia tweet after the outlets were removed.

Figures VII and VIII together paint a consistent picture. On both the supply and the demand side, we only find effects of the ban for the most prolific pre-ban suppliers, who both created pro-Russia content on at least four of the eight days leading up to the ban and who were in the top 10% of suppliers with respect to the engagement their average pro-Russia tweet had received.

Consistent with the channels outlined in our conceptual framework in Subsection 2.3, there are two main explanations for the ban affecting the most prolific pre-ban suppliers. First, as mentioned in the previous section, we show in Appendix Figure B.3 that the pre-ban most active secondary suppliers also had the closest connections with the banned outlets. These users likely relied the most on Russia Today and Sputnik to create their own content. Therefore, the ban affected them through increased production costs of pro-Russia content. Second, the effects we find are also consistent with a signaling effect of the ban. The most prolific pre-ban suppliers could have interpreted the ban as a warning that pro-Russia content would face stronger moderation, feared that another ban wave would target them, and reduced the number of their pro-Russia tweets. Less prolific suppliers of pro-Russia content faced both of these concerns to a lesser degree, explaining the muted effect we find for them. We can think of similar explanations for the drop in demand as proxied by engagement as for the decrease in supply. The high-activity and top-engagement suppliers reduced their content and benefited less from social media's cascading effects as other users shared their content less. Alternatively the explanation may come from the demand side: the followers of the biggest secondary accounts may have seen the ban as a signal that pro-Russia content is tainted or risky. As a result, they were less willing to like, retweet, or reply—so engagement fell, echoing the 'warning effect' we argued producers felt.

These results, though limited in data and scope, offer a cautionary perspective on the effectiveness of the ban. While we find an overall decrease in the amounts of pro-Russia content in the previous sections, we detect limited effects of the outlet's removal on other suppliers of pro-Russia tweets. Remarkably, in this very short-run setting, we find no evidence for attempts to substitute the banned outlets as predicted by the theoretical literature modeling news consumption as seeking out belief-consistent information (Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006).

4.6 Russia Today and Sputnik as Agenda Setters

The evidence we presented so far points to Russia Today and Sputnik not just being peripheral actors, but central players in inspiring and sustaining the spread of pro-Russia narratives in Europe. Their disappearance from the media sphere marked a decisive shift in the ecosystem that produced such content. In the final part of our analysis on observational data, we try to understand the mechanism behind the observed effects of the ban: Why were Russia Today and Sputnik such effective instruments of pro-Russia propaganda? And what made them so central to the machinery of content production?

To answer these questions, we examine the impact of the ban on the number and type of topics contained in pro-Russia tweets. Besides assigning the labels pro-Russia, anti-Russia, or neutral content, our unsupervised learning approach extracts up to five of the most salient topics mentioned in each tweet. This topic information allows us to measure topic diversity in pro-Russia content before and after the ban. If Russia Today and Sputnik were indeed key sources of inspiration for framing and producing pro-Russia narratives, we would expect their removal to reduce topic diversity. In other words, without centralized

TABLE IV
IMPACT OF THE BAN ON PRO-RUSSIA CONTENT TOPICS

Dependent variable	**Topics rate**	**# top 5 outlets' topics per tweet**
	Coeff./SE/p-value	
	(1)	(2)
Ban x EU	-0.071	-0.020
	(0.013)	(0.002)
	[0.000]	[0.000]
User FEs	✓	✓
Date FEs	✓	✓
Conditional on posting about war	✓	✓
Pre-ban outcome avg. for treated	1.051	0.085
Approx. percentage change	-6.73	-1.98
Observations	322199	322199

Notes: The table examines the effect of the ban on pro-Russia topics, classified by our LLM-based pipeline, which identifies up to five topics per tweet. We present estimates of two-way fixed effects difference-in-differences regressions on the user-day panel dataset from February 22nd to March 15th. Column 1 reports estimates of the ban's effect on the daily average number of topics in a user's pro-Russia tweets using Equation 1. Column 2 shows the effect of the ban on the average number of topics in a user's tweets that overlap with the outlets' top five topics on the same day, estimated via Equation 1. We compute the approximate percentage change in Column 1 as the change from the pre-ban average and in Column 2 as $e^{\beta} - 1$. All specifications include user and day fixed effects, and standard errors are clustered at the user level.

guidance, users might struggle to maintain the same breadth of content, narrowing the scope of their messaging. Table IV presents this analysis.

For this set of results, we return to our full sample again. In Column 1, we report the effect of the ban on the average number of topics per pro-Russia tweet. In Column 2, we analyze the overlap between the topics in users' tweets and the top 5 topics pushed by the banned outlets on the specific day. In both cases, the results point to a consistent interpretation. The ban disrupted the supply chain of pro-Russia content by removing key sources of inspiration; among European users, the average number of topics in pro-Russia tweets falls by roughly 6%. Similarly, we observe a decline in the share of topics in user-generated content that overlap with the top five themes most actively promoted by the outlets on a given day.

Taken together, these findings provide suggestive evidence that Russia Today and Sputnik functioned as agenda setters central to a network of users, signaling which themes to emphasize and how to frame them. Once these key actors were removed, the remaining pro-Russia network found itself left without a clear guide to structure its narratives or articulate a coherent worldview.

5 Robustness Checks, Placebo Tests, and Additional Results

In this section, we summarize additional output from our observational data analysis, focusing on robustness checks, placebo tests, and supplementary findings (reported in detail in the appendices). These exercises validate the main results presented above and provide further context for their interpretation.

Robustness Checks: Our first robustness check evaluates the reliability of the classification pipeline used throughout the paper, as we mentioned earlier on in the paper. While there is substantial evidence supporting the accuracy and effectiveness of Large Language Model (LLM) architectures for text analysis, Appendix F presents a validation exercise. We construct an alternative measure of pro-Russia content – which we call *media slant* – we (i) compare it to our main classification and (ii) replicate the main results of the paper using this alternative.

We build this alternative measure following the approaches of Gennaro and Ash (2023) and Gentzkow and Shapiro (2011). Specifically, we collect tweets from representatives of the Russian and Ukrainian governments active on Twitter. We vectorize and average these tweets to generate two reference poles representing the "average opinion" of each side. We then vectorize tweets from users in our sample and compute their relative distances to these poles. Tweets closer to the Russian pole than the Ukrainian one are classified as pro-Russia. The resulting measure aligns closely with our original classification, and reproducing our core analysis shows qualitatively similar results.

Second, we address the potential influence of bots on our results. There is no single optimal definition or method for detecting bots in data of the type we use; however, in Appendix G we apply several definitions inspired by the Computer Science literature to identify users that could plausibly be bots. Excluding these users from the analysis does not alter our main results. Moreover, if Russian bots had attempted to counteract the effects of the ban, such activity would likely bias our estimates toward zero, making our results a lower bound of the ban's impact. In addition, Appendix H replicates our results after excluding users whose accounts were created only after the ban took effect, with no substantive change in the findings.

Third, we ask whether a correlated shock might threaten our identification strategy. Our difference-in-difference strategies rely on the parallel trends assumption but might also be confounded by an unobserved shock coinciding with the ban that affected EU users differently than non-EU users in their stance towards Russia more broadly. While we are not aware of any high-frequency survey data that would allow us to test this directly, we turn to the demand for Russian culture as a plausible proxy for attitudes towards Russia. We create a dataset of relative Google Search volume for a consistent set of nine representative Russian cultural figures across the larger countries in our sample – Germany, France, Italy, and the UK. In Appendix K, we test whether we detect a change in the demand for Russian culture coinciding with the ban, measured by the relative amount of Google searches for important Russian cultural figures. We find no statistically significant changes in search behavior in the EU compared to non-EU users. While suggestive, we take this as supporting evidence that the effects observed in our main analysis are directly attributable to the ban and not an artifact of a correlated shock that shifted perceptions towards Russia differently in the treated and control group.

A final robustness check concerns our model specification. Throughout most of the analysis, we rely on Poisson-Pseudo Maximum Likelihood (PPML) models, which recent literature identifies as a preferred approach for count data (see, for example, Chen and Roth 2024). Nevertheless, in Appendix I we re-estimate our main specifications using OLS instead of PPML. While the magnitudes differ, the results point in the same direction, supporting the robustness of our findings.

Placebo Tests: We run several placebo tests to validate our empirical findings. First, we examine whether the ban represents a breaking point in the production of pro-Russia content among users in non-EU countries. Figure IVa provides descriptive evidence that, for EU users, pro-Russia content begins a strong and steady decline immediately after the ban. When we reproduce the same specification for non-EU countries – shown in Appendix Figure C.1 – no such pattern emerges. This provides a visual check that the observed decline in the EU is not part of a broader global trend unrelated to the ban.

Second, we complement the evidence above with an additional placebo test. A potential concern in our setting is that the differential impact of the ban documented above might instead reflect some underlying, idiosyncratic factors unrelated to the ban itself. To test this, we reproduce our main event study specification but assign a fictitious ban date set one week earlier than the actual implementation, in Appendix Figure D.1. If our results were driven by such confounding factors, we might observe a similar differential effect at this earlier date. Instead, the fictitious ban shows no effect on pro-Russia content in EU versus non-EU countries, providing reassurance that the actual breaking point in our analysis is indeed caused by the ban itself.

Third, we run a placebo to explore whether the ban also had an impact on the anti-Russia content in our sample. Users spreading anti-Russia content can be seen as the counterpart of those spreading pro-Russia narratives: they are motivated to promote their worldview, and they do so largely independently of what the opposing side does. This implies that the ban, which specifically targeted pro-Russia outlets, should have little or no effect on the production of anti-Russia content. Testing this provides a useful placebo: if we were to find a decline in anti-Russia content after the ban, it would suggest the presence of broader confounding factors rather than an effect specific to pro-Russia speech. Appendix Figure D.2 confirms this expectation, showing no systematic change in anti-Russia content following the ban.

Additional Results: In the final part of this section, we explore whether there is any evidence of an institutional reaction to the ban. 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. Thus, once lost, these major amplifiers, it is plausible that Russian institutional resources, attention, and effort were redirected elsewhere. In Appendix J we explore this by showing the impact of the ban on another important source of information from Russia: TASS, the state news agency.

In the appendix, we compare the activity and engagement of TASS with those of Russia Today and Sputnik, before and after the ban. If institutional attention had shifted to TASS, we would expect to see an increase in its output. Likewise, if users had redirected their demand to this alternative source, we would expect higher engagement with TASS's pro-Russia content. Yet we find no evidence of such changes. A plausible explanation is that TASS and the banned outlets played different roles in the information ecosystem: TASS primarily functioned as a wire service, focused on short news updates rather than sustained propaganda framing consistently spread on social media.

6 Cost of censorship

Our analysis suggests that censorship can be an effective tool in a democratic context. Yet, when democracies rely on emergency powers and restrictive measures to control speech, they risk eroding the very norms that form the foundation of the system (Linz 1978). Historical experience offers clear warnings. In Weimar Germany, the executive used Article 48 of the constitution (emergency powers) to curtail civil liberties – including bans on political assemblies, press censorship, and party bans – especially targeting communists and eventually the Nazi Party (Evans 2004). For example, radio access was restricted for Hitler and the NSDAP during the presidential campaign in 1932 (Adena et al. 2020). Once in power, the Nazis could build on the precedence created for the quick and comprehensive *Gleichschaltung* of the media, effectively putting all media under direct control of the Nazi government.

In this section, we examine the consequences of using censorship in a democratic context. Specifically, we ask: Was there any cost brought by the EU's decision to ban Russia Today and Sputnik on its own democratic legitimacy? To address this question, we use a simple online experiment to examine how perceptions of core democratic institutions change when the use of censorship by the democratic regime is made salient. Below we provide an overview of the design and the results of this experiment.

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.[12] 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, we ask all participants to express their general satisfaction with how democracy works in the EU and then to read some 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.[13] 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 agreement with the statement "The EU does not guarantee the

[12] 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.
[13] See Appendix L.2 for the full information briefs and all screens used in the online experiment.

independence of media" (reverse coded). We use a 7-point Likert scale to measure agreement with the statements. 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 that mask the purpose of the study, further measuring trust in national government, national parliament, the European Parliament, and the European Commission, as well as satisfaction with how democracy works in the EU. The survey concludes with questions on sociodemographic characteristics. Appendix L.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 L.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 V 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 of Russian state-backed media outlets, although the effect is statistically significant at the 10% level in a two-sided t-test only when we include all sociodemographic controls and baseline satisfaction with democracy.[14] When examining the two items separately, we find a clear negative treatment effect on the freedom of speech item (Table V, Column 2), statistically significant at the 5% level when including all relevant controls. The media independence item also suggests a negative treatment effect, but it is not statistically significant (Table V, Column 3).[15] For all other items on democratic norms and trust in institutions, we find only small and statistically insignificant effects (Appendix L.2 and L.3).

In the final part of this section, we explore some heterogeneous effects of the informational treatment. Although we did not preregister an analysis by political orientation, it is informative to explore whether the treatment effect varies across the political spectrum. Political stance may shape sensitivity to restrictions on freedom of speech, with some individuals more tolerant of curbing civil liberties in the name of security. Those holding more extreme positions may also be more open to radical policy measures. Historical patterns add further nuance: in many EU countries, left-wing movements have traditionally maintained closer ties to Russia, while recent years have seen the rise of right-wing groups that portray Russia as a political model. Differences in perceived threat, trust in EU institutions, and media consumption habits

[14] 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 0.095 of a standard deviation in the freedom of speech index. Our sample is designed to detect a minimum effect size of 0.2 of a standard deviation in a two-sided test. Given the effect size we observe, we appear to be underpowered to clearly identify the effect.

[15] This item was reverse-coded, and we are potentially underpowered to detect an effect.

TABLE V

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	-0.163	-0.075
	(0.072)	(0.079)	(0.100)
	[0.095]	[0.040]	[0.452]
Language FEs	✓	✓	✓
Baseline satisfaction with democracy	✓	✓	✓
Individual characteristics	✓	✓	✓
Mean dep. var.	0.484	0.932	0.031
Approx. percentage change	-11.33	-17.46	-241.48
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 6 for details). Column 1 uses the pre-registered *Freespeech Index* as the dependent variable. It is computed as the average of approval to the *Freedom of speech* and *Media Independence* items recorded separately in columns 2 and 3. All items use a 7-point Likert approval scale to indicate whether the European Union protects the corresponding construct. *Media Independence* was reverse coded. 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.

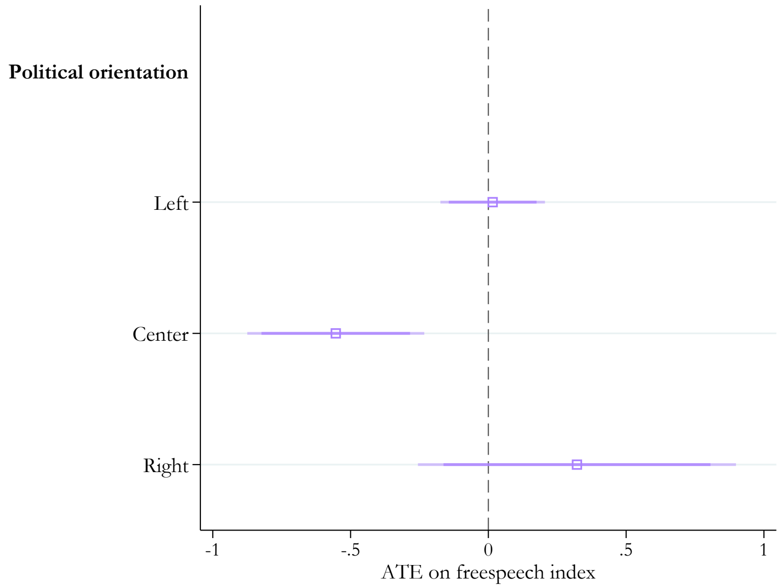
may therefore produce systematic variation in how participants respond to information about the media ban.

Figure IX 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-leaning (N=87) do not show any negative response to the treatment.[16] If anything, the point estimate for right-leaning respondents goes in the opposite direction.

In sum, the results from the experiment suggest that using censorship in a democratic context can have a meaningful cost by eroding trust in the democratic institutions themselves. In this simple experiment, the effect appears to be largely domain-specific and concentrated on perceptions of freedom of speech, while not extending to broader measures of trust in the democratic order. Nevertheless, the findings clearly indicate the potential for a dynamic tradeoff between the short-term effectiveness of top-down regulatory measures to curb misinformation and the long-term erosion of trust in democratic norms. This tradeoff should be a core concern when applying strong restrictive measures such as censorship in a democratic context. Moreover, our results on the heterogeneity of the effect by political orientation provide suggestive evidence that the use of censorship in a democratic context might be particularly

[16] 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 L.2).

FIGURE IX
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 6 for details) by the political orientation of the respondent. Dependent variable is the *Freespeech Index* computed as the average of approval to the *Freedom of speech* and *Media Independence* items. All items use a 7-point Likert approval scale to indicate whether the European Union protects the corresponding construct. *Media Independence* was reverse coded. 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.

costly by eroding the trust of the democratic center – arguably the core constituency upholding the democratic order. At the fringes of the political spectrum, and among those already dissatisfied with democracy, the use of more authoritarian measures to curb misinformation might not be perceived as negative.

7 Conclusion

In this paper, we study the effects and consequences of using censorship in a democratic context. We leverage the European Commission's decision to ban all broadcasting activity of Russia Today and Sputnik to provide causal evidence that the ban was effective in reducing pro-Russian content online. We find the most pronounced effect among users directly connected to the outlets prior to the ban and among highly active pre-ban suppliers of pro-Russian content. We find no evidence of substitution by other suppliers or a shift of attention towards other suppliers. This indicates that the ban increased the cost of production of pro-Russian content that could not be compensated for in the short run. We also illustrate a key mechanism behind the effectiveness of the ban by pointing to the role of the banned outlets as agenda setters. We document that overlap with topics pushed by the outlets on a given day is reduced among users affected by the ban, highlighting their central role in providing key narratives that suppliers of pro-Russian content were relying on.

We want to highlight a few key limitations of this study. First, we are only able to observe a narrow time window around the ban and hence cannot make any assessment of the longer-term dynamics of the effects. We are not able to study persistence or account for any substitution in the behavior of institutions, suppliers, and users that fall outside the short initial two-week time period of the ban. Second, we only observe the conversation on one platform, Twitter. There is scope for future research to assess patterns of cross-platform substitution that have been documented to matter in other settings (Rizzi 2024). While cross-platform substitution is a concern in principle, we think it is less relevant in our setting as the ban was not specific to one platform but affected all broadcasting activity of Russia Today and Sputnik in the European Union.

Finally, we provide evidence of the potential cost of using censorship in a democratic context. While our empirical analysis of the ban's effect on the narratives spread on Twitter about the Russian-Ukrainian conflict suggests that the ban might have been effective in reducing the volume of pro-Russian content as intended by the European Commission, our online experiment shows that the use of censorship can undermine trust in core features of the democratic order itself. Specifically, we show that increasing the salience of the use of censorship by the EU reduces trust in the notion of the EU upholding the principles of freedom of speech and press. This result illustrates a dynamic tradeoff that democratic governments face, built around the *Paradox of Tolerance* (Popper 1945): democracies might need to retain the right to deny tolerance to counter the threat of misinformation and spread of foreign propaganda trying to undermine the open society. On the other hand, using harsh regulatory measures to counter such threats might be an effective tool in the short run but undermine the core pillars on which the democratic order itself rests in the long run.

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.

Adena, Maja et al. (2020). "Bombs, Broadcasts and Resistance: Allied Intervention and Domestic Opposition to the Nazi Regime During World War II". In: *SSRN Electronic Journal*.

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.*

Arechar, Antonio A. et al. (June 29, 2023). "Understanding and combatting misinformation across 16 countries on six continents". In: *Nature Human Behaviour* 7.9, pp. 1502–1513.

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.

Brown, Tom B. et al. (2020). *Language Models are Few-Shot Learners*. Version Number: 4.

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, Georgy Egorov, and Stefano Fiorin (Nov. 1, 2020). "From Extreme to Mainstream: The Erosion of Social Norms". In: *American Economic Review* 110.11, pp. 3522–3548.

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.

Ershov, Daniel and Juan S Morales (Aug. 22, 2024). “Sharing News Left and Right: Frictions and Misinformation on Twitter”. In: *The Economic Journal* 134.662, pp. 2391–2417.

Evans, Richard J. (2004). “The Coming of the Third Reich”. In: *History Review* 50, pp. 12–17.

Fazio, Lisa (Feb. 10, 2020). “Pausing to consider why a headline is true or false can help reduce the sharing of false news”. In: *Harvard Kennedy School Misinformation Review*.

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.

Germano, Fabrizio, Vicenç Gómez, and Gaël Le Mens (2019). “The Few-Get-Richer: A Surprising Consequence of Popularity-Based Rankings?” In: *The World Wide Web Conference*, pp. 2764–2770.

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.

Henry, Emeric, Ekaterina Zhuravskaya, and Sergei Guriev (2022). "Checking and Sharing Alt-Facts". In: *American Economic Journal: Economic Policy* 14.3, pp. 55–86.

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*.

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*. Publisher: arXiv Version Number: 3.

Nyhan, Brendan (Aug. 1, 2020). "Facts and Myths about Misperceptions". In: *Journal of Economic Perspectives* 34.3, pp. 220–236.

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.

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.

Skafle, Ingjerd et al. (Aug. 4, 2022). “Misinformation About COVID-19 Vaccines on Social Media: Rapid Review”. In: *Journal of Medical Internet Research* 24.8, e37367.

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.

Wang, Wenhui et al. (2020). “MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers”. In: *Advances in neural information processing systems 33*, pp. 5776–5788. arXiv: 2002.10957[cs].

West, Robert (2020). “Calibration of Google Trends Time Series”. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. CIKM ’20: The 29th ACM International Conference on Information and Knowledge Management. Virtual Event Ireland: ACM, pp. 2257–2260.

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

Yaqub, Waheeb et al. (Apr. 21, 2020). “Effects of Credibility Indicators on Social Media News Sharing Intent”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20: CHI Conference on Human Factors in Computing Systems. Honolulu HI USA: ACM, pp. 1–14.

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°°

Appendix

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

**University of Bern, Wyss Academy at the University of Bern, e-mail: *matteogrigoletto@gmail.com*

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

Appendix

Table of Contents

A Materials and Methods

In this Appendix, we offer additional information on the materials and methods used in our analysis. Specifically, we detail the process of data retrieval in Section A.1, describe our method for user localization in Section A.2, and explain the classification of tweets in Section A.3.

A.1 Collection of Tweets

This section provides insights into the process of extraction and processing of data from Twitter. 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.[17] All the tweets we downloaded were posted between January 24$^{\text{th}}$, 2022, and April 4$^{\text{th}}$, 2022. Using the extracted data we create two datasets on data collected from Twitter: (1) a dataset of Ukrainian and Russian government-associated accounts and (2) a sample of tweets posted by users involved in the discussion about the unfolding conflict and invasion of Ukraine.

The first dataset consists of tweets posted by accounts affiliated with the Russian and Ukrainian governments. We provide a full list of the accounts in Table F.1 in Appendix F, 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.

The second 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*. 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 a.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.

A.2 Geolocation of Users

To create the dataset comprising tweets from general Twitter users, we employ a geo-location process for user identification. It is crucial to remember that Twitter data acquired via the API does not automatically include geo-tag information. This means that while our query seeks tweets related to the conflict, these tweets could originate from users all around the world. To construct a dataset specifically from users in the EU, we proceed as follows. Our initial download resulted in data from 1,942,979 users. Utilizing the geo-location method outlined in Gehring and Grigoletto (2025), we then identify users located in our target countries: Austria, France, Germany, Ireland, Italy, Switzerland, and the UK, narrowing the

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

group to 146,633 users. For these users, we subsequently download all tweets matching our query: *russ* OR ukrain* OR nato OR otan*. This process yields a dataset of 677,780 tweets (original tweets, replies, and retweets with no more than 140 characters, as we extend on in the next Subsection) posted between February 22$^{\text{nd}}$ and March 15$^{\text{th}}$. For detailed information on the geolocation methodology, please see Gehring and Grigoletto (2023).

A.3 AI 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 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)" }"*

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.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 10% 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 provides descriptive statistics for the main sample of tweets used in the analysis. It is extensively commented and discussed in the Paper Section 3. Table B.3 presents descriptive statistics for users involved in our analysis. This includes a total of 146,633 users who were selected as part of our sample; more on the selection process in Appendix A. 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. In the table, we also show that roughly 5% of users were directly in contact with the banned outlets. Additionally, roughly 39% of users are located in the countries of the control group, the United Kingdom and Switzerland, while the others are in the countries of the treatment group, Austria, France, Germany, Ireland, and Italy.

TABLE B.2
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	2.2	3	1.7	0	5
RT vs. Sputnik					
Share of tweets by Russia Today	.61	1	.49	0	1
Observations	4,032				

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	25
Connection to Russia Today and Sputnik					
Ever connected to the outlets	.64	1	.48	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 3.2 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 3.2. 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.

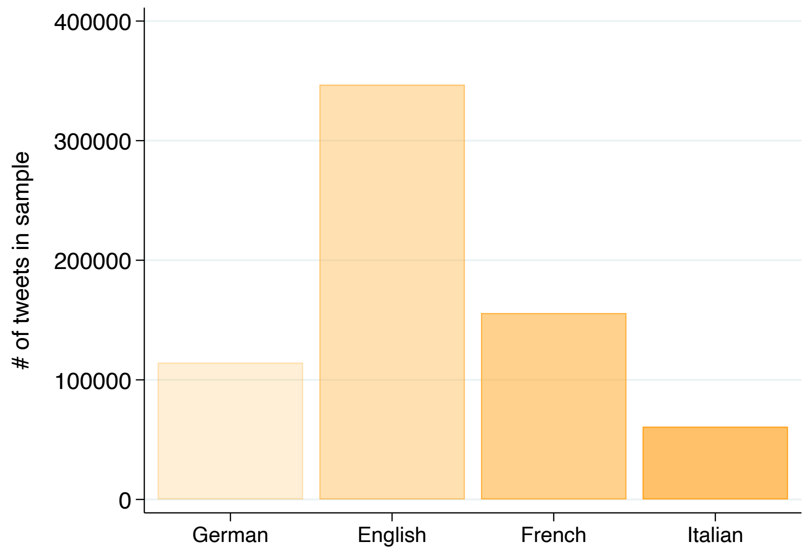
TABLE B.3
USER LEVEL DESCRIPTIVES

	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					
Ever Connected to the Outlets	.42	0	.49	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 3.2 posted between February 22nd, 2022, and March 15th, 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 most spoken languages in Europe: English, German, French, and Italian. As the figure shows, English language tweets prevail, making up almost as much as the remaining languages pulled together. 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

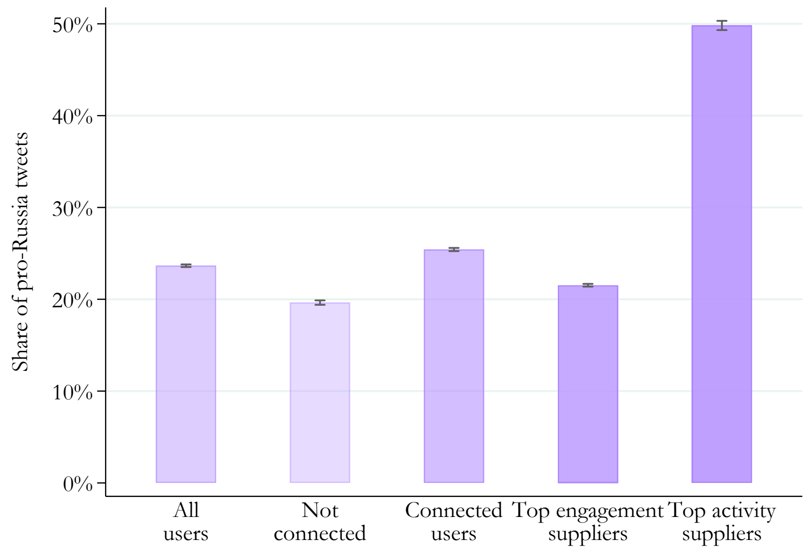


Notes: The figure shows the distribution of tweets in our sample by language of the original post. We translate all non-English tweets to English before putting them into the classification pipeline. The Table B.2 provides additional insights into the data used in the analysis.

Figure B.2 provides additional insights into the share of pro-Russia tweets in our sample. The figure reports the pre-ban share of pro-Russia content for five samples of users. From left to right, the bars represent: (i) all users in the corpus; (ii) users who never interacted directly with either outlet; (iii) users who interacted with at least one outlet before the ban; (iv) top-engagement secondary suppliers, defined as the top 10% of users by audience engagement on their pro-Russia tweets; and (v) high-activity secondary suppliers, i.e the accounts that produced the largest absolute number of pro-Russia tweets before the ban. As an important reminder, these are all the sub-samples used in our analysis; the second and third in the list are mutually exclusive, while top engagement and top activity users may encompass users from both the connected and non-connected.

The descriptive pattern is consistent with the supplier role of Russia Today and Sputnik. Users who never interacted with the outlets display the lowest incidence of pro-Russia content, whereas the share noticeably increases among users who did interact with them. Top-engagement secondary suppliers show no elevated shares of pro-Russia content while high-activity secondary suppliers stand out; this group posts the highest proportion of pro-Russia content, approaching 50% of its total output. These differences underscore the importance of distinguishing user types when evaluating the ban's downstream effects.

FIGURE B.2
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 sub-sample of users included in our analysis. We include observations from February 22$^{\text{nd}}$ to March 1$^{\text{st}}$, 2022. The sub-samples, shown from left to right in the figure, are as follows: (1) all users; (2) non-connected users (those who never retweeted or replied to the banned outlets); (3) connected users (those who replied to or retweeted the outlets at least once prior to the ban); (4) top-engagement secondary suppliers (users in the top 5% of engagement on their pro-Russia tweets); and (5) high-activity secondary suppliers (those who posted pro-Russia content on at least four of the eight days before the ban).

Table B.4 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*, *negotiations*, 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, and on the day after, becomes the single most cited one. 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. After the EU ban takes effect on 1$^{\text{st}}$ March, the topic *sanctions* rises into — and consistently then remains in — the daily top five, signaling a sustained 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.4
TOP 5 OUTLET TOPICS BY DAY

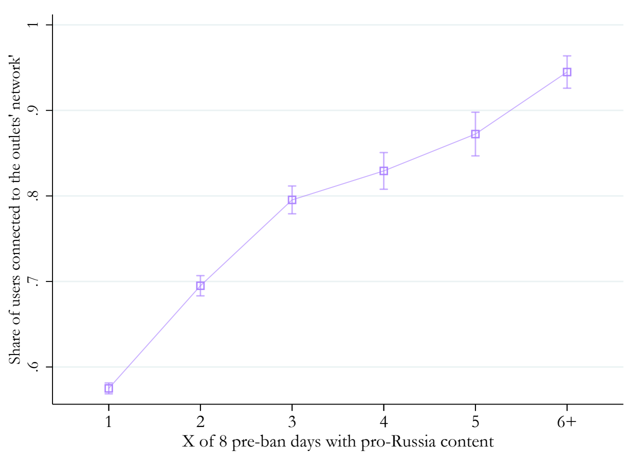
Date	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
2022–02–16	media bias	syria	military drills	belarus	military exercises
2022–02–17	nato expansion	security guarantees	us	military drills	crimea
2022–02–18	donetsk	evacuation	donbass	media bias	explosion
2022–02–19	donetsk	nato expansion	evacuation	dpr	biden
2022–02–20	donbass	diplomacy	macron	shelling	donetsk
2022–02–21	donbass	donetsk	recognition	nato expansion	lugansk
2022–02–22	recognition	sanctions	donbass	lpr	donetsk
2022–02–23	sanctions	donbass	dpr	nord stream 2	recognition
2022–02–24	donbass	military operation	sanctions	lugansk	donetsk
2022–02–25	sanctions	media bias	military operation	diplomacy	negotiations
2022–02–26	sanctions	media	media bias	swift	conflict
2022–02–27	media bias	sanctions	eu sanctions	belarus	china
2022–02–28	sanctions	media bias	negotiations	belarus	refugees
2022–03–01	media bias	sanctions	un	conflict	nuclear weapons
2022–03–02	sanctions	media bias	diplomacy	belarus	negotiations
2022–03–03	sanctions	media bias	media	censorship	diplomacy
2022–03–04	sanctions	nato expansion	media bias	nuclear power	military
2022–03–05	media bias	sanctions	mariupol	media	ceasefire
2022–03–06	donbass	conflict	sanctions	zelensky	media bias
2022–03–07	media bias	talks	diplomacy	moscow	refugees
2022–03–08	sanctions	china	military operation	gas prices	russian oil
2022–03–09	sanctions	us	chernobyl	economic war	china
2022–03–10	sanctions	diplomacy	media bias	turkey	russian military
2022–03–11	sanctions	media bias	violence	meta	hate speech
2022–03–12	sanctions	violence	media bias	russian forces	ukraine war
2022–03–13	media bias	sanctions	conflict	china	us sanctions
2022–03–14	sanctions	china	donetsk	media bias	us
2022–03–15	sanctions	china	russian military	mariupol	diplomacy

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 IV is based on these topics.

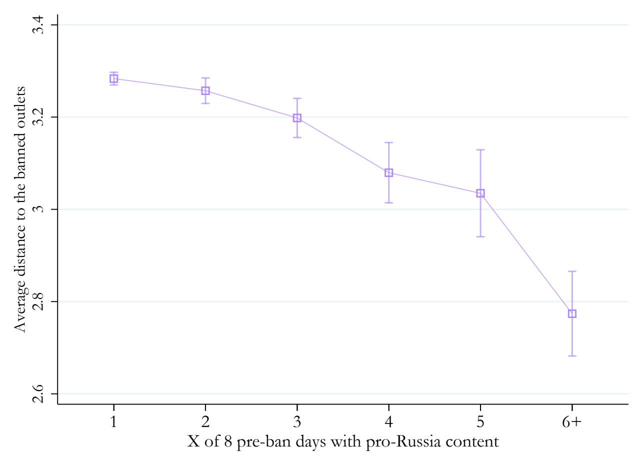
Figures B.3 and B.4 show the connection between pre-ban activity of secondary suppliers with being part of the networks of Russia Today and Sputnik, and the engagement secondary suppliers' pro-Russia content managed to attract before the ban. We document the co-movement of activity, network membership, 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 secondary suppliers' characteristics by forming mutually exclusive combinations of user groups for our heterogeneity analysis.

FIGURE B.3
SECONDARY SUPPLIERS' ROLES IN THE OUTLETS' NETWORK BY LEVEL OF PRE-BAN ACTIVITY

(a) SECONDARY SUPPLIERS' CONNECTIONS TO RT AND SPUTNIK

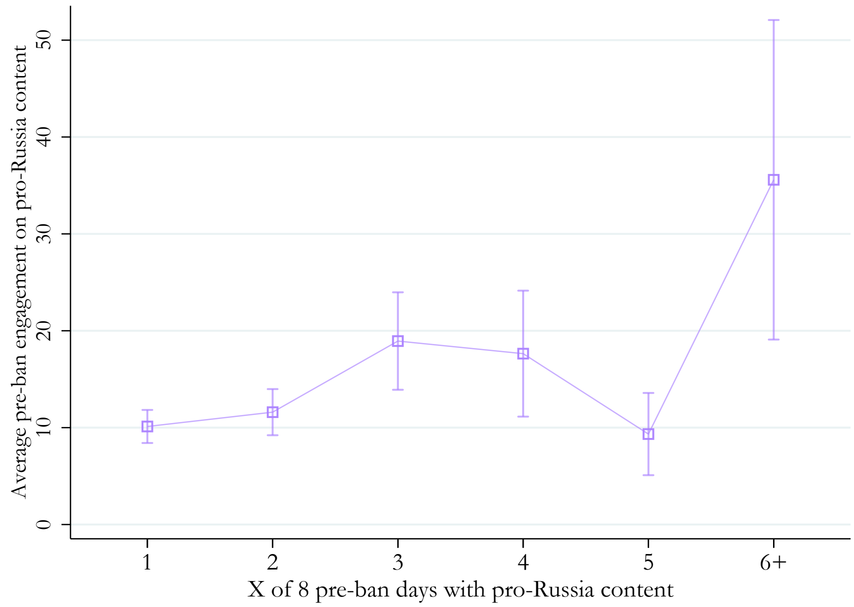


(b) SECONDARY SUPPLIERS' DISTANCE TO RT AND SPUTNIK



Notes: The figure visualizes correlations of secondary suppliers' 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.

FIGURE B.4
SECONDARY SUPPLIERS' PRE-BAN ENGAGEMENT BY LEVEL OF PRE-BAN ACTIVITY



Notes: The figure plots the average pre-ban engagement of pro-Russia tweets by secondary suppliers relative to the number of pre-ban days with pro-Russia activity.

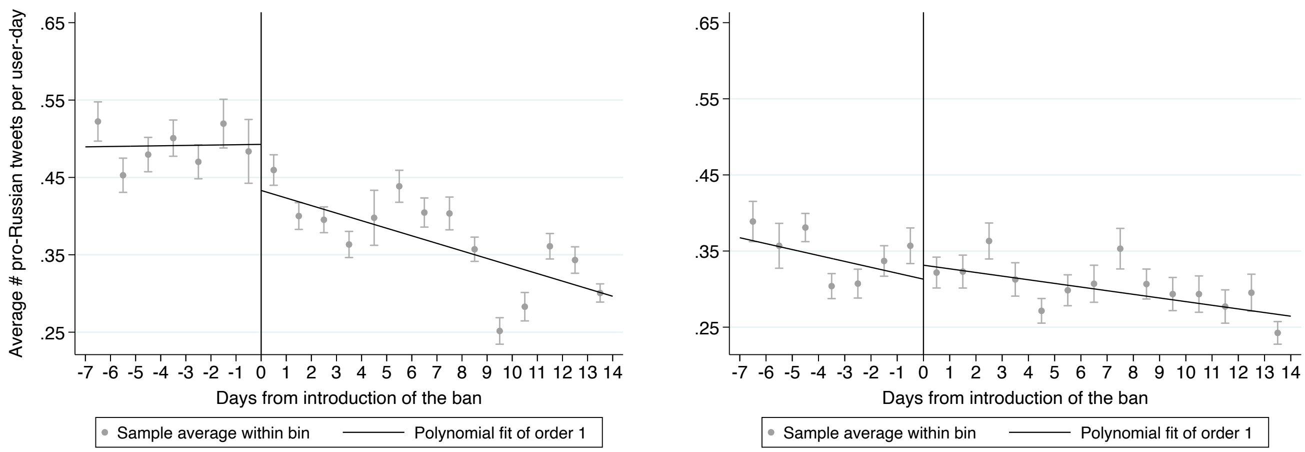
C Additional Results

In this section, we provide additional results complementing the output we report in the main paper. We start from Figure C.1 that expands on the descriptive evidence reported in the main text. The left-hand panel reproduces the Paper Figure IVa, for ease of reference: it plots the daily average number of pro-Russia tweets posted by users located in the EU, our treatment group. The plot is obtained with an event study framework using time as the running variable and indicating, as a breaking point, with a vertical line, the day of the enforcement of the ban. A first-degree polynomial is fitted on either side of the break. As we highlight in the main text, the timeseries exhibits an immediate downward shift at the policy date, followed by a persistent decline, providing a visual confirmation that the supply of pro-Russia content fell, once Russia Today and Sputnik were removed from the EU information space.

The right-hand panel applies the same specification to the non-EU countries in our sample (the United Kingdom and Switzerland). Here, the trajectory of pro-Russia tweets is essentially flat, with only a mild, gradual decline over time and no clear change at the breaking point. This divergence in slopes and break magnitudes reinforces the identification strategy: it suggests that, absent the ban, EU trends would have mirrored the stable pattern observed in the control group, lending credibility to the natural-experiment interpretation of our results.

FIGURE C.1
EVENT STUDY: PRO-RUSSIA CONTENT

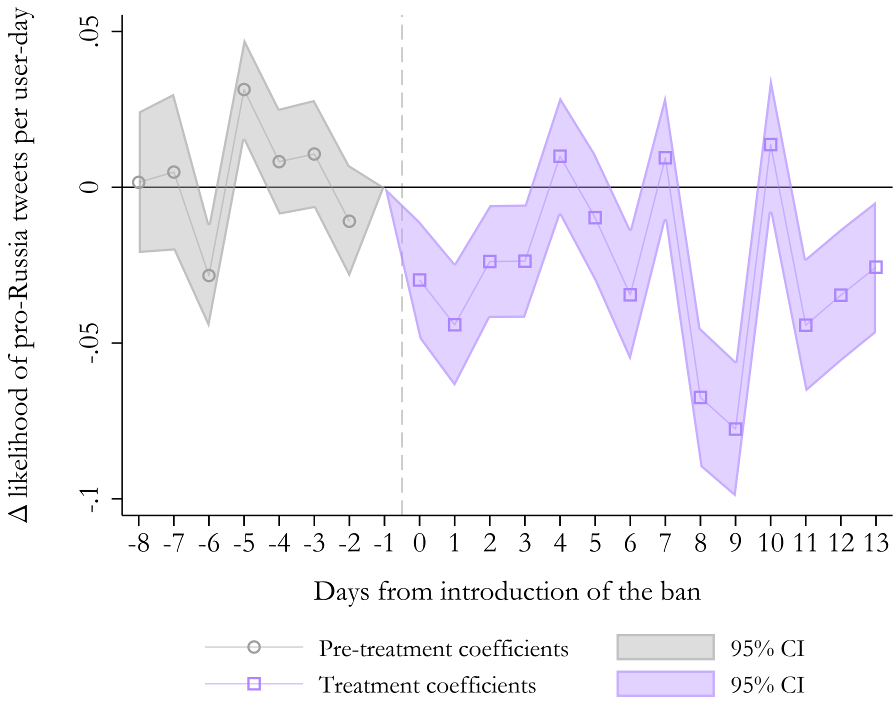
(a) EU (b) NON-EU



Notes: The figures show Regression Discontinuity Design (RDD) analysis, where the running variable is time. The coefficients trace the evolution of the absolute number of pro-Russia tweets posted by users located in EU countries around the introduction of the ban. The vertical axis shows the daily average, and we apply a linear fit on either side of the cutoff. The horizontal axis reports the number of days relative to the ban (day 0), ranging from 7 days before to 14 days after its implementation. Shaded areas denote 95% confidence intervals. Panel (a) shows results for the EU countries, already reported in the paper, while panel (b) shows the same for non-EU countries. We include only observations between February 22nd and March 15th, to ensure a large enough sample size for each day. Back to the Paper Figure IVa, showing the event study for EU countries.

In the next figure, we provide additional evidence of the effect of the ban on European users vs non-EU users. Figure C.2 reproduces the exact same specification of the Paper Figure V, while using a different dependent variable, the likelihood of posting pro-Russia content in a given day, instead of the absolute number of pro-Russia tweets posted. The aim of this additional investigation is the following. If, on the one hand, the absolute number of pro-Russia tweets captures an important measure of the effect of the ban, which we can think of as the extensive margin, this may still be driven by very few high-level producers of content, suddenly changing their activity. The likelihood of posting captures a complementary measure, which we think of as the intensive margin, and can only be driven by the choice of posting or not posting pro-Russia content. Reassuringly, the figure confirms the main results of the paper. Despite presenting a noisier outline, overall, the results corroborate what we see in the main paper, with the ban having a negative impact also on the likelihood of posting pro-Russia on a given day.

FIGURE C.2
DAILY EVENT STUDY: IMPACT OF THE BAN ON LIKELIHOOD
OF POSTING PRO-RUSSIA CONTENT

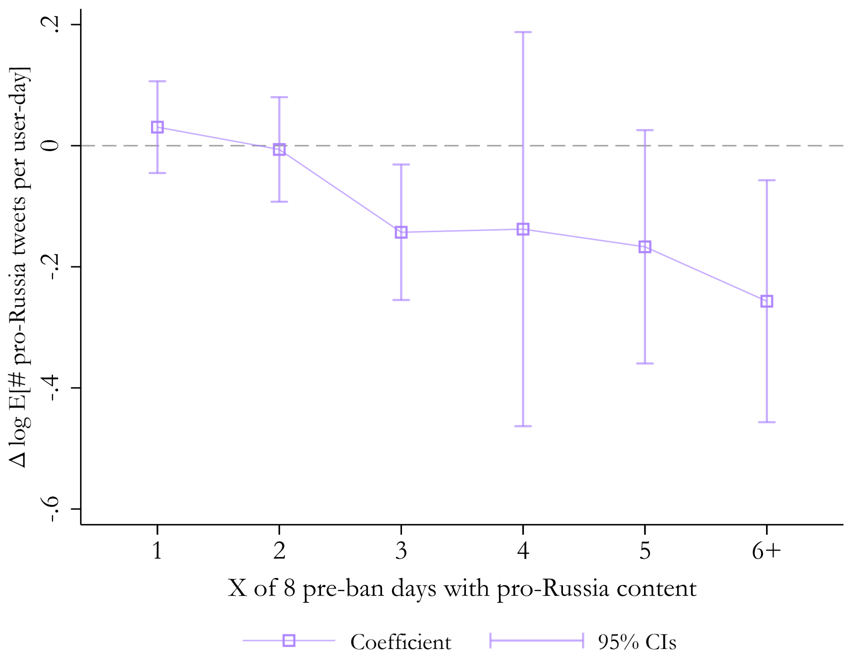


Notes: The figure displays coefficients and 95% confidence intervals from a linear probability regression model estimated in an event study framework. We use the user-day panel dataset and include only user-day observations where the user was active in our sample as described in Section 3.2. We include only observations between February 22nd and March 15th, to ensure a large enough sample size for each day. The dependent variable is a dichotomous variable equal to 1 if the user posted pro-Russia content on a given day, 0 otherwise. The model includes user and day fixed effects, and 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. Back to the Paper Figure V, showing the same model for the absolute number of pro-Russia tweets.

Figure C.3 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 secondary suppliers, 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. 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 E.2.

The figure reveals a clear pattern: the ban's negative effect grows in magnitude with the number of days a secondary supplier was active in producing pro-Russia content before its implementation, with the largest decline among the most active subgroup and even a reversal of sign for those active on only one day. While we cannot disentangle the exact mechanism behind this pattern, we offer some explanation in Appendix Figures B.3 and B.4. Panel B.3a shows that the share of secondary suppliers

FIGURE C.3
IMPACT ON SECONDARY SUPPLIERS OF PRO-RUSSIA CONTENT
BY LEVEL OF PRE-BAN ACTIVITY



Notes: The figure displays coefficients and 95% confidence intervals from estimating Equation 4 while exploring the impact of the ban on secondary suppliers of pro-Russia content. Users are defined as secondary suppliers if they posted at least one pro-Russia tweet in the eight days before the ban. We run six separate regressions, one for each subgroup of users defined by their 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. In Appendix Table E.2, we report complete information on the regressions, including the number of observations and the percentage change for the corresponding difference-in-difference estimation by transforming the coefficients β as $e^{\beta} - 1$. All reported regressions include user and day fixed effects, and standard errors are clustered at the user level.

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.3b, we additionally show that the average distance of connected secondary suppliers is decreasing in activity. Therefore, the most active users were also the most dependent on the banned outlets for content and the most aware of the ban and its signaling effect, explaining the largest decrease in pro-Russian content. To shed further light on the secondary suppliers and their activity, Figure B.4 plots the average engagement – the sum of retweets, replies, and likes – of pro-Russian content posted by secondary suppliers separated by their activity. This descriptive output reveals a further dimension between the activity of secondary suppliers and their pro-Russia content. The average pre-ban engagement on pro-Russian content of secondary suppliers is increasing in activity. Based on the documented overlap between the secondary suppliers' characteristics, we form mutually exclusive combinatorial subgroups for the heterogeneity analysis in the main part of the paper.

Finally, in Table C.1 we give an indication of the implied reduction in pro-Russia tweets per day of our main difference-in-difference estimate reported in Table I. We calculate the ”missing” pro-Russia tweets by calculating the implied reduction in pro-Russia tweets per user-days via the estimated percentage change of the difference-in-difference estimation and the pre-ban number of pro-Russia tweets per user-day in the European Union. Finally, we multiply this reduction by the number of pre-ban users in the European Union in our sample. The result of this back-of-the-envelope calculation is a reduction of 3224 pro-Russia tweets per day due to the ban.

TABLE C.1
BACK-OF-THE-ENVELOPE QUANTIFICATION

	Description	**Value**
Step 1	Estimated % change in pro-Russian tweets per user-day from Table I	-10.92%
Step 2	Pre-ban number of pro-Russia tweets per user-day in European Union	.487
Step 3	Implied reduction in number of pro-Russia tweets per user-day	-.053
Step 4	Implied reduction in number of pro-Russia tweets per day (= Step 3 · Number of pre-ban users in European Union [= 60,830])	-3224

Notes: The table summarizes the steps and the intermediate results of a back-of-the-envelope calculation of the main effect of the ban on Russia Today and Sputnik. We quantify the effect as the implied decrease in the number of pro-Russia tweets per day from the EU.

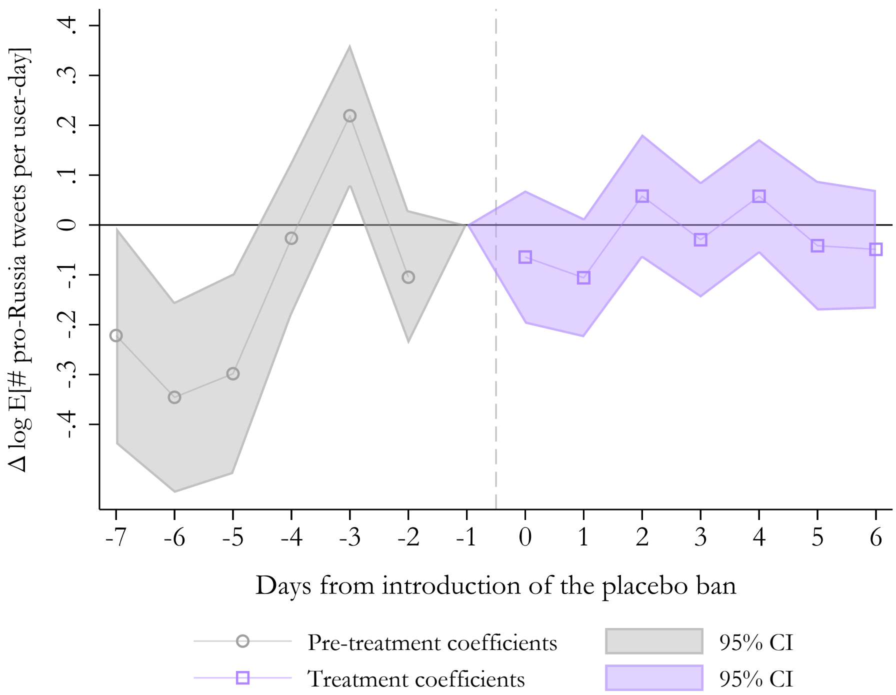
D Placebo Checks

In this section, we present two additional checks that probe the internal validity of our natural experiment. We begin with a placebo test on the timing of the intervention. 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 D.1 replicates the specification used in Figure V, but shifts the break point to a fictitious ban date one week 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 accounts for the pre-trends effects visible in the pre-trends. Overall, the evidence supports the validity of our identification strategy.

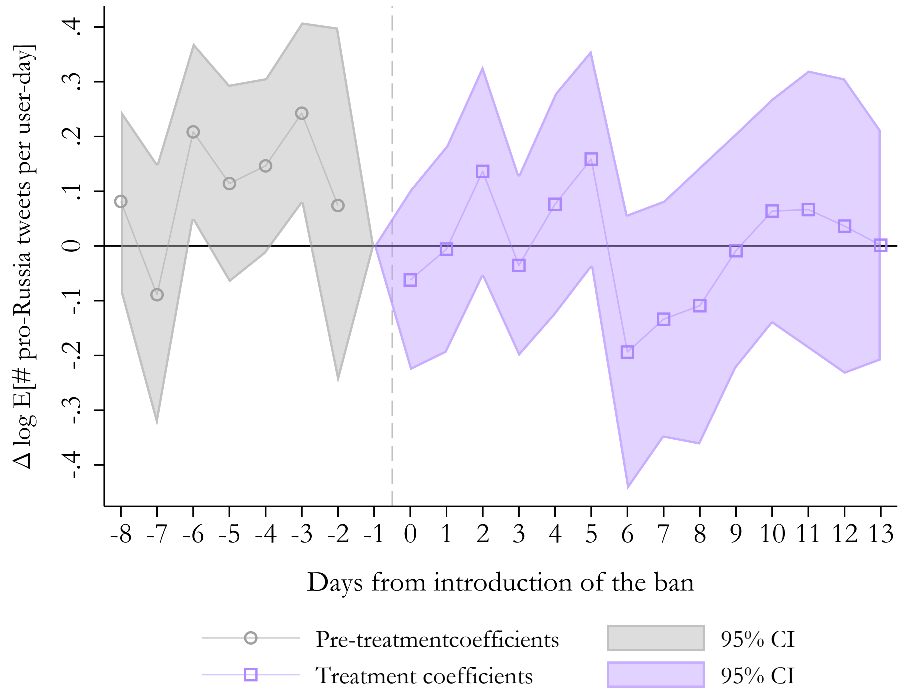
The second placebo exercise, reported in Figure D.2, investigates whether the ban influenced anti-Russia content. As a reminder, our classification pipeline assigns every tweet a value label: pro-Russia when the text aligns with Kremlin narratives, neutral, or anti-Russia when it opposes them. The core 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 D.2 shows that the ban leaves the volume of anti-Russia tweets essentially unchanged.

FIGURE D.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 Equation 4. We use a placebo event, shifting the implementation to one week earlier than the actual ban, i.e. from March 2$^{\text{nd}}$ to February 25$^{\text{th}}$, 2022. 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 3.2. We include only observations between 16$^{\text{nd}}$ 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 ban on a user's pro-Russia tweet activity, conditional on the user being active that 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. In Table I, we report the percentage change for the corresponding difference-in-difference estimation by transforming the coefficients β as $e^{\beta} - 1$. The reported regression includes user and day fixed effects, and standard errors are clustered at the user level. The omitted day is February 24$^{\text{nd}}$, 2022, the day immediately preceding the introduction of the placebo ban. The vertical dashed line marks the date of the policy intervention. Back to the Paper Section 5 exploring robustness checks and additional results.

FIGURE D.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 Equation 4. 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 3.2. 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 anti-Russia tweets posted per user per day; thus, the coefficients capture the daily average effect of the ban on a user's anti-Russia tweet activity, conditional on the user being active that 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. 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 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. Back to the Paper Section 5 exploring robustness checks and additional results.

E Additional Details on Regression Models

This section provides additional details for the main results of the paper. In particular, we provide the underlying models for all the outputs of the paper we present in figures and not in tables. Table E.1 shows the underlying models of the Paper Figure VI; similarly, Table E.2 for the Paper Figure VII, Table E.3 for the Paper Figure VII.

TABLE E.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 x EU	-0.182	-0.042
	(0.058)	(0.031)
	[0.002]	[0.181]
User FEs	✓	✓
Date FEs	✓	✓
Conditional on Posting About War	✓	✓
Pre-Ban Outcome Avg. for Treated	0.948	0.438
Approx. Percentage Change	-16.64	-4.11
Observations	51132	97632

Notes: The table displays the results from estimating Equation 4 corresponding to the paper Figure VI. We present results from two separate regressions that share an identical specification but differ in the user samples analyzed. Both regressions use the user-day panel dataset and include only user-day observations where the user was active in our sample as described in Section 3.2. 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 follows: e^{β} -1. All regressions include user and day fixed effects, and standard errors are clustered at the user level.

TABLE E.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 x EU	0.127	0.040	-0.150	-0.115	-0.195	-0.273
	(0.037)	(0.038)	(0.050)	(0.125)	(0.090)	(0.077)
	[0.001]	[0.303]	[0.003]	[0.355]	[0.030]	[0.000]
User FEs	✓	✓	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓	✓	✓
Conditional on Posting About War	✓	✓	✓	✓	✓	✓
Pre-Ban Outcome Avg. for Treated	0.145	0.329	0.563	0.953	1.235	1.929
Approx. Percentage Change	13.56	4.04	-13.91	-10.87	-17.74	-23.87
Observations	73799	36990	20600	12688	8497	9037

Notes: The table displays the results from estimating Equation 4 corresponding to Appendix Figure C.3 while exploring the impact of the ban on secondary suppliers of pro-Russia content. Users are defined as secondary suppliers if they posted at least one pro-Russia tweet in the eight days before the ban. We run six separate regressions, one for each subgroup of users defined by their 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 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. The approximate percentage change is computed as follows: e^{β} -1. All reported regressions include user and day fixed effects, and standard errors are clustered at the user level.

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

Dependent Variable	**Total No. of Pro-Russia Tweets**			
Samples	**Low Act., and Regular Engag.**	**Low Act., and Top Engag.**	**High Act., and Regular Engag.**	**High Act., and Top Engag.**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban x EU	0.037	0.091	0.000	-0.318
	(0.030)	(0.078)	(0.087)	(0.160)
	[0.221]	[0.247]	[0.999]	[0.047]
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Conditional on posting about war	✓	✓	✓	✓
Pre-ban outcome avg. for treated	0.243	0.475	1.172	2.292
Approx. percentage change	3.79	9.50	0.02	-27.25
Observations	73916	11095	7394	5066

Notes: The table displays the results from estimating Equation 4 while investigating the heterogeneous effect of the ban on secondary suppliers corresponding to the paper Figure VII. All regressions use the user-day panel dataset and include only user-day observations where the user was active in our sample as described in Section 3.2. We include only observations between February 22nd and March 15th, to ensure a large enough sample size for each day. The figure divides secondary suppliers by their activity before the ban and the amount of engagement their pro-Russia content received before the ban. For activity, we separate suppliers into those who produced pro-Russia content in three or fewer days before the ban (low activity) and those who produced pro-Russia content in more than three days before the ban (high activity). With respect to engagement, we distinguish between regular and top engagement. The top engagement includes users whose content received the top 5% number of retweets, likes, and replies. Regular engagement captures anything else. 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 follows: e$^{\beta}$ -1. All reported regressions include user and day fixed effects, and standard errors are clustered at the user level.

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

Dependent Variable	**Avg. engagement per tweet**			
Samples	**Low Act., and Regular Engag.**	**Low Act., and Top Engag.**	**High Act., and Regular Engag.**	**High Act., and Top Engag.**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Ban x EU	0.160 (0.257) [0.532]	0.178 (0.296) [0.547]	0.005 (0.312) [0.986]	-0.421 (0.251) [0.094]
User FEs	✓	✓	✓	✓
Date FEs	✓	✓	✓	✓
Conditional on posting about war	✓	✓	✓	✓
Pre-ban outcome avg. for treated	1.945	99.401	1.334	43.579
Approx. percentage change	17.38	19.51	0.53	-34.36
Observations	23769	4571	5125	6427

Notes: The table displays the results from estimating Equation 4 while investigating the heterogeneous effect of the ban on the engagement, which the pro-Russia content of different groups of secondary suppliers receives, corresponding to the paper Figure VIII. All regressions use the user-day panel dataset and include only user-day observations where the user was active in our sample as described in Section 3.2. 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 secondary suppliers by their activity before the ban and the amount of engagement their pro-Russia content received before the ban. For activity, we separate suppliers into those who produced pro-Russia content in three or fewer days before the ban (low activity) and those who produced pro-Russia content in more than three days before the ban (high activity). With respect to engagement, we distinguish between regular and top engagement. The top engagement includes users whose content received the top 5% number of retweets, likes, and replies. Regular engagement captures anything else. The dependent variable is the engagement – the sum of likes, retweets, and replies – secondary suppliers achieve on their average pro-Russia tweet per user per day; coefficients therefore capture the daily average effect of the ban on engagement, 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 follows: e^{β} -1. All reported regressions include user and day fixed effects, and standard errors are clustered at the user level.

F Robustness Check: 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 (Wang et al. 2020), 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 F.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.

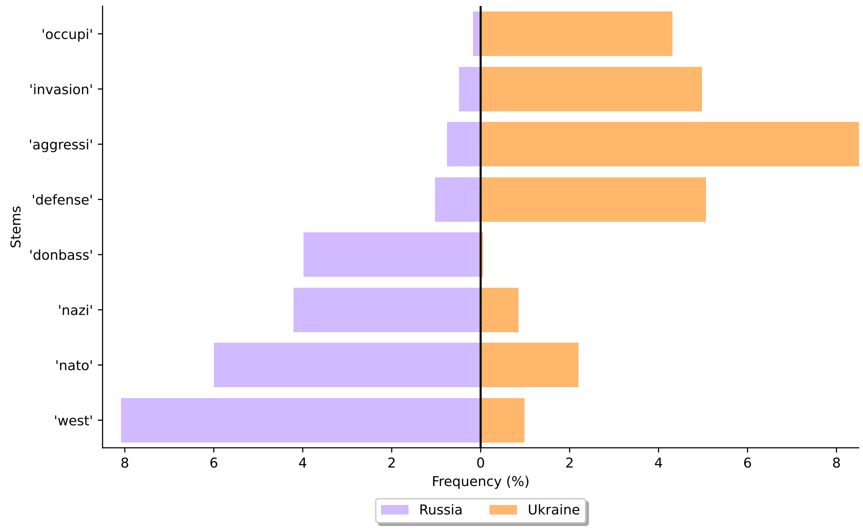
TABLE F.1 ACCOUNTS OF THE RUSSIAN AND UKRAINIAN GOVERNMENTS' REPRESENTATIVES

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	Colonel General Oleksandr Syrskyi	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 F.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 purple 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 F.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 purple, 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 3.2) with sentence-t5-xl and use Equation 5 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:

$$Y = \frac{sim(d, R) + b}{sim(d, U) + b} - 1, \tag{5}$$

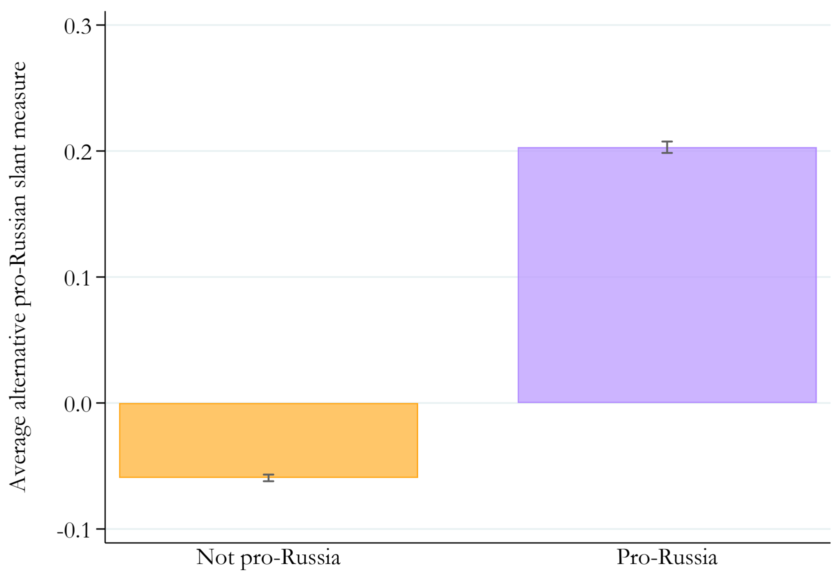
where d denotes a vector representing the input text, R and U are vectors representing the two poles, b is a smoothing parameter set to 1, and sim refers to the cosine similarity. We subtract 1 to center the ratio around zero. Positive values indicate tweets more similar to the Russian pole. Tweets with 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[18]. The pole ratio for each user tweet from day t is then computed based on the two corresponding daily

[18] 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 F.2 plots the average continuous pro-Russia slant measure divided by the binary pro-Russia slant measure 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.

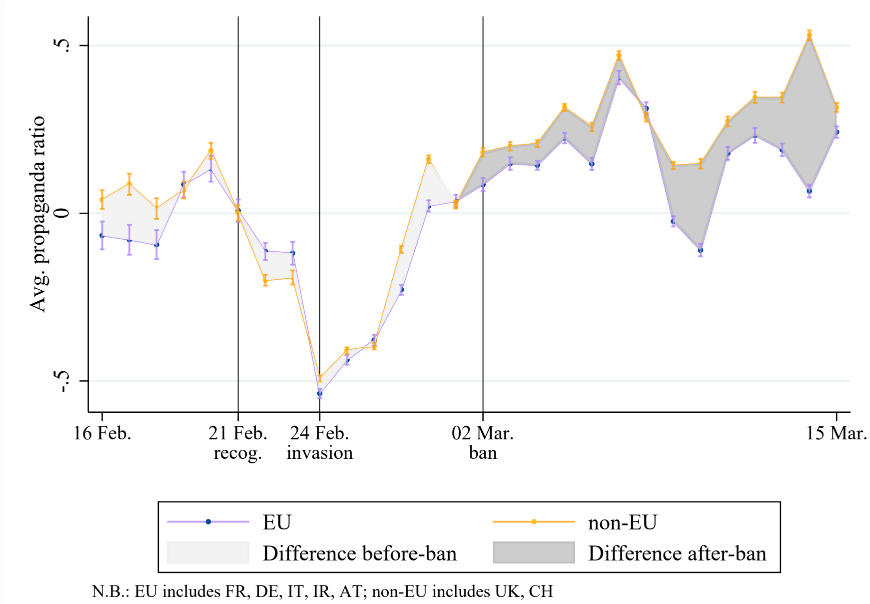
FIGURE F.2 COMPARISON BETWEEN BINARY AND CONTINUOUS PRO-RUSSIA SLANT



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

In Figure F.3, we present the daily average measure of media slant in both EU and non-EU countries as part of our analysis. The graph reveals several notable trends. In the days marking the onset of the invasion, both EU and non-EU regions reached a minimum in average slant, suggesting a widespread initial reaction leaning towards the Ukrainian pole. Subsequently, there is a pronounced and consistent shift towards the Russian pole, underscoring the European Commission's concern that the conflict was being waged not only on the ground but also online, with the EU particularly targeted by Russian propaganda efforts. Furthermore, the movement of average slant in EU and non-EU countries shows similar trends before the ban's implementation, after which a distinct divergence is observed.

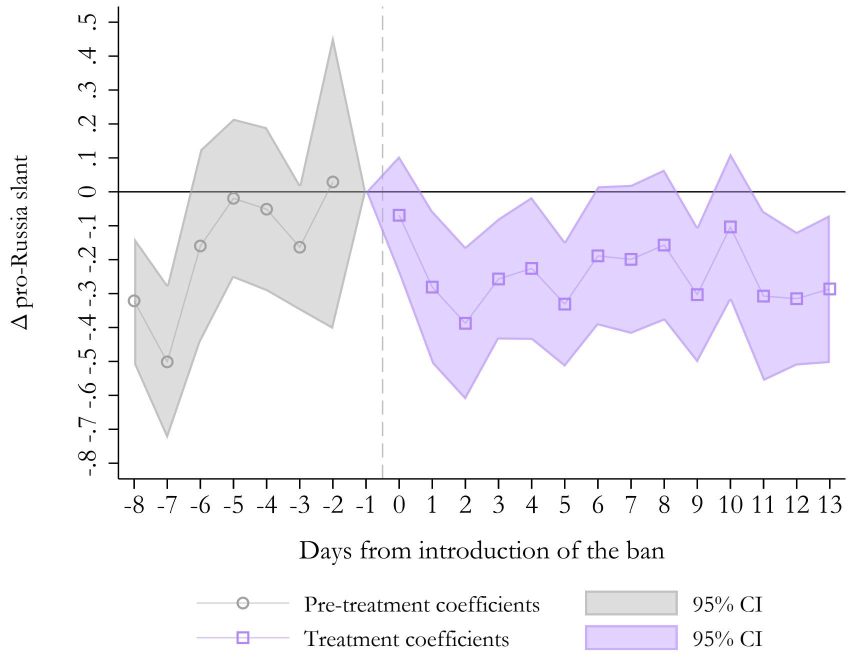
FIGURE F.3 TIME-SERIES OF OUR SLANT MEASURE: DAILY AVERAGES



Notes: The figure shows the daily averages of our media slant measurement, in the time frame of our analysis, between February 19th, 2022, to March 15th, 2022. The measure is normalized to have a mean of 0 and a standard deviation of 1. When positive, the measure indicates content closer to the Russian pole, and when negative, it indicates content closer to the Ukrainian pole. In purple, we show the daily averages in the EU countries in our study, Austria, France, Germany, Ireland, and Italy, and in orange, the daily averages in non-EU countries, the United Kingdom and Switzerland. In grey, the difference between the two averages. We indicate the following important dates: 21st Feb. for the official recognition by Putin of the Donetsk People's Republic and the Luhansk People's Republic, 24th Feb. as the beginning of the war and 2nd Mar. as the beginning of the ban.

Finally, in Figure F.4 and Table F.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 F.4 plots the results of estimating the event study specification in Equation 3 for the continuous slant measure. We find a similar negative shift in pro-Russia content as in Figure V. 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 F.2 reproduces the decrease in pro-Russia content presented in Table I. 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 Table II, also holds for the alternative slant measure.

FIGURE F.4 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 from estimating Equation 3. 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 3.2. 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. The reported regression includes user and day fixed effects, and 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.

TABLE F.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 x EU	-0.142 (0.044) [0.001]	-0.205 (0.059) [0.001]	-0.033 (0.059) [0.578]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	✓
Pre-ban outcome avg. for treated	-0.274	-0.236	-0.339
Approx. percentage change	-51.84	-86.60	-9.69
Observations	322199	202671	119528

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 Equation 1. 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 3.2. 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 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 reproduces the results of Columns 1 and 2 of the Paper Table I, using the alternative continuous measure of pro-Russia slant instead of the discrete measure of pro-Russia tweets of the main parts. Column 2 shows the effect for users connected to the outlets, reproducing the results of Columns 1 and 2 of the Paper Table II Panel A. Column 3 presents the effect for non-connected users, reproducing Columns 1 and 2 of the Paper Table II Panel B. All specifications include user and day fixed effects, and standard errors are clustered at the user level.

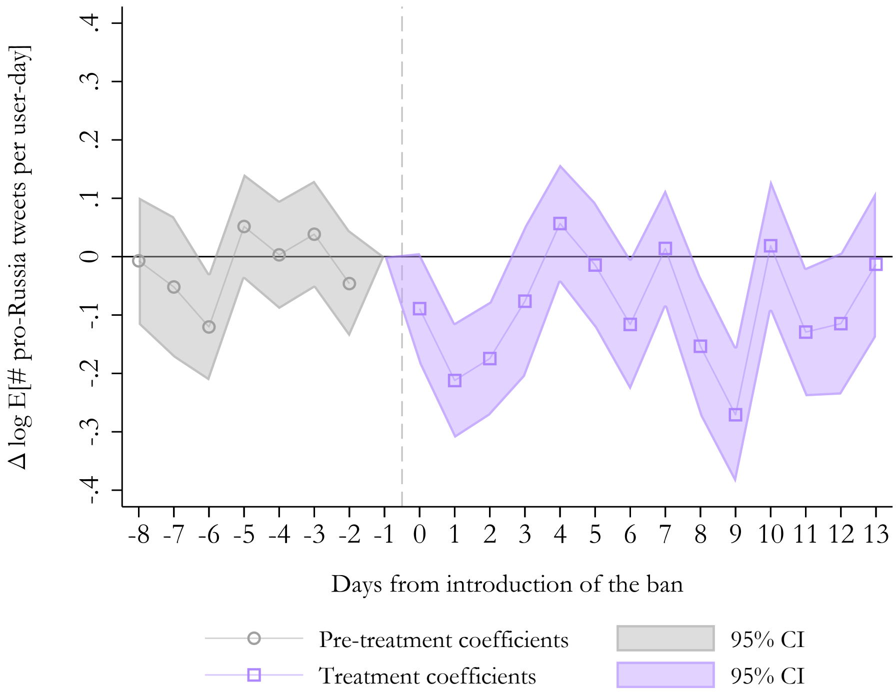
G Robustness Check: Potential Bots

In this Appendix, 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 target an ad hoc subgroup of suspiciously active accounts: users who produced more than 80 pro-Russia tweets in the eight days preceding the ban. Applying these criteria leads to the exclusion of 2,585 users.

In Figure G.1 and Table G.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 anything, we would expect undetected Russian bots to attempt to counteract the effect of the ban. Thus, the effects we find may represent a lower bound of the actual impact of the ban.

FIGURE G.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 Equation 4. It reproduces the results of the Paper Figure V. 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 3.2, and users that are not potential bots. 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 on average 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 a user's pro-Russia tweet activity, conditional on the user being active that 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 March 1st, 2022, the day immediately preceding the introduction of the ban. The vertical dashed line marks the date of the policy intervention.

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

Dependent variable	**# tweets pro-Russia**	**P(Tweet: Pro-Russia)**	**P(Tweet: About war)**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Ban x EU	-0.072 (0.020) [0.000]	-0.027 (0.003) [0.000]	-0.007 (0.001) [0.000]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	
Pre-ban outcome avg. for treated	0.472	0.324	0.160
Approx. percentage change	-6.95	-8.42	-4.17
Observations	209494	313384	3268628

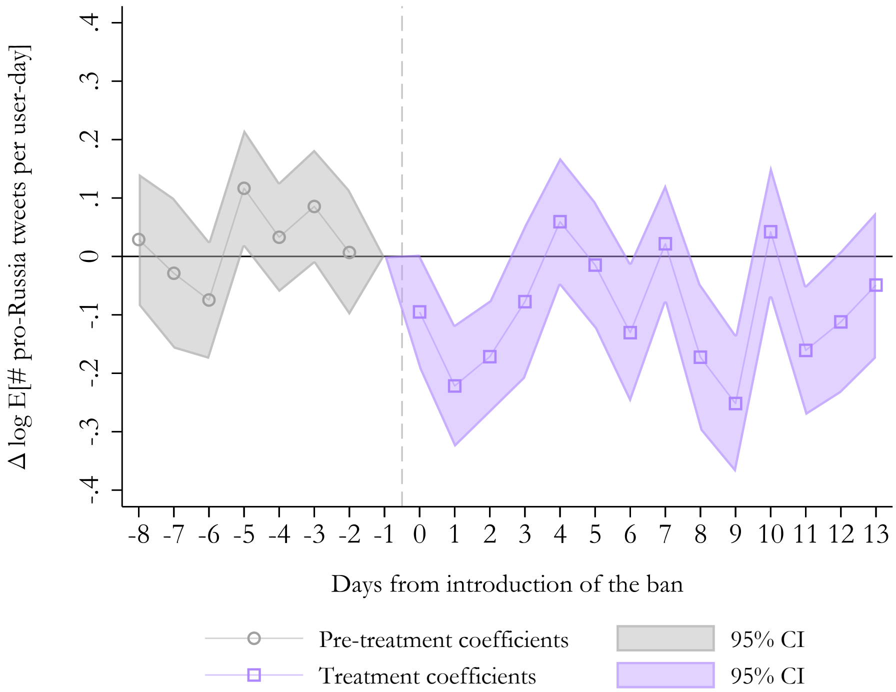
Notes: The table presents the results from two-way fixed effects difference-in-difference regressions analyzing the impact of the ban on users' posting behavior. It reproduces the results of the Paper Table I. All models use the user-day panel dataset and include observations from February 22^{nd} to March 15^{th}, 2022. We exclude potential bots, 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 on average more than 80 pro-Russia tweets in the eight days before the ban. Column 1 reports estimates on the impact on the total number of pro-Russia tweets via Equation 2, conditional on tweeting about the war. Column 2 shows estimates of posting pro-Russia content conditional on tweeting about the war using Equation 1. Column 3 reports estimates of posting any war-related content, again via Equation 1. We compute the approximate percentage change in Column 1 as $e^{\beta} - 1$, and in Columns 2 and 3 as the change from the pre-ban average. All specifications include user and day fixed effects, and standard errors are clustered at the user level.

H Robustness Check: Accounts Created After the Ban

In this Appendix, 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 H.1 and Table H.1.

FIGURE H.1

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 Equation 4. It reproduces the results of the Paper Figure V. 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 3.2, excluding those accounts that were created after the ban. 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; thus, the coefficients capture the daily average effect of the ban on a user's pro-Russia tweet activity, conditional on the user being active that 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 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.

TABLE H.1
IMPACT OF THE BAN ON PRO-RUSSIA AND OVERALL WAR-RELATED CONTENT EXCLUDING ACCOUNTS CREATED POST-BAN

Dependent variable	**# tweets pro-Russia**	**P(Tweet: Pro-Russia)**	**P(Tweet: About war)**
		Coeff./SE/p-value	
	(1)	(2)	(3)
Ban in EU vs. Ban in Non-EU	-0.116	-0.027	-0.007
	(0.028)	(0.003)	(0.001)
	[0.000]	[0.000]	[0.000]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	
Pre-ban outcome avg. for treated	0.487	0.323	0.162
Approx. percentage change	-10.93	-8.49	-4.51
Observations	215329	321547	3315422

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. It reproduces the results of the Paper Table I. All models use the user-day panel dataset and include observations from February 22$^{\text{nd}}$ to March 15$^{\text{th}}$, 2022. We exclude accounts that were created after the ban. Column 1 reports estimates on the impact on the total number of pro-Russia tweets via Equation 2, conditional on tweeting about the war. Column 2 shows estimates of posting pro-Russia content conditional on tweeting about the war using Equation 1. Column 3 reports estimates of posting any war-related content, again via Equation 1. We compute the approximate percentage change in Column 1 as $e^{\beta} - 1$, and in Columns 2 and 3 as the change from the pre-ban average. All specifications include user and day fixed effects, and standard errors are clustered at the user level.

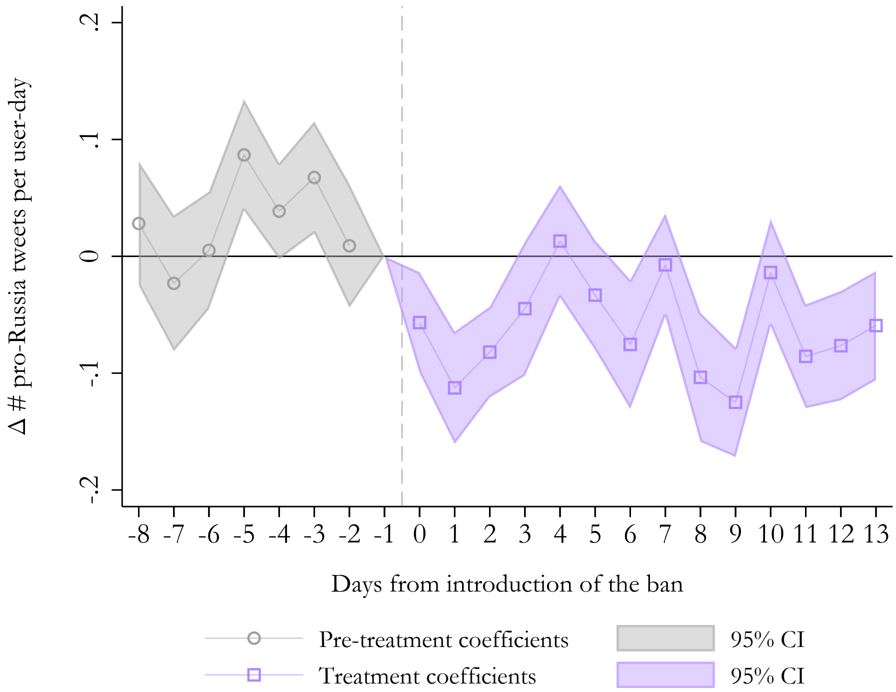
I Robustness Check: PPML vs. OLS

In this Appendix, we provide additional details on our econometric choices and consider potential alternatives. Recall that our main analysis relies on three dependent variables: (i) the number of pro-Russia tweets produced by users, conditional on posting about the war; (ii) the probability that, given a user posts about the war, the tweet is pro-Russia; and (iii) the probability of tweeting about the war at all. For the second and third outcomes, we estimate simple OLS probability models. The first outcome, however, poses greater challenges and requires more careful consideration.

In this Appendix, we provide additional details on our econometric choices and consider potential alternatives. Recall that our main analysis relies on three dependent variables: (i) the number of pro-Russia tweets produced by users, conditional on posting about the war; (ii) the probability that, given a user posts about the war, the tweet is pro-Russia; and (iii) the probability of tweeting about the war at all. For the second and third outcomes, we estimate simple OLS probability models. The first outcome, however, poses greater challenges and requires 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: even if the size of the coefficients changes, the direction of the results should remain consistent. Figure I.1, Table I.1, and Figure I.2 reproduce the most important results of our analysis, substituting PPML models with simple OLS models. Reassuringly, the results present no qualitative change from what we show in the main analysis.

FIGURE I.1
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 Equation 3. 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 3.2. 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 a user's pro-Russia tweet activity, conditional on the user being active that day. The reported regression includes user and day fixed effects, and 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.

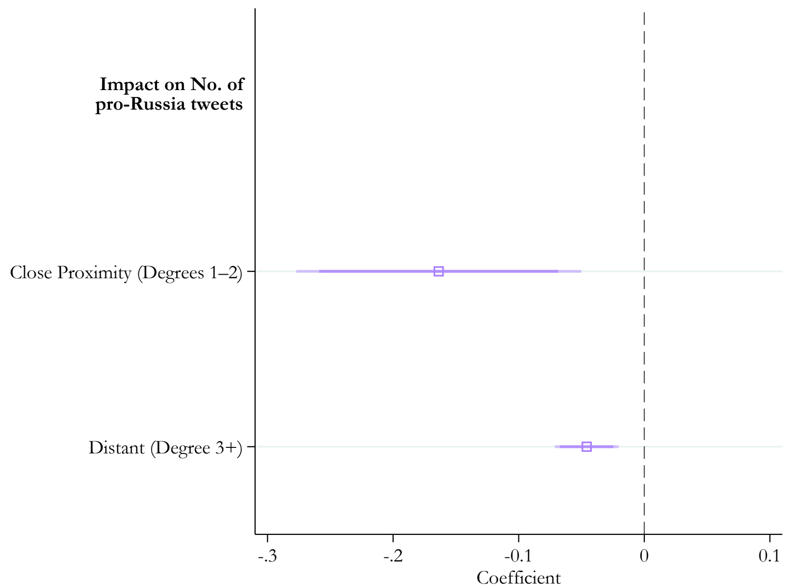
TABLE I.1

IMPACT OF THE BAN ON PRO-RUSSIA AND OVERALL WAR-RELATED CONTENT

Dependent variable	**# tweets pro-Russia**	**P(Tweet: Pro-Russia)**	**P(Tweet: About war)**
		Coeff./SE/p-value	
	(1)	(2)	(3)
Ban x EU	-0.091	-0.027	-0.007
	(0.015)	(0.003)	(0.001)
	[0.000]	[0.000]	[0.000]
User FEs	✓	✓	✓
Date FEs	✓	✓	✓
Conditional on posting about war	✓	✓	
Pre-ban outcome avg. for treated	0.487	0.323	0.161
Approx. percentage change	-18.75	-8.49	-4.46
Observations	322199	322199	3325498

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. It reproduces the results of the Paper Table I. All models use the user-day panel dataset and include observations from February 22^{nd} to March 15^{th}, 2022. Column 1 reports estimates on the impact on the total number of pro-Russia tweets via Equation 1, conditional on tweeting about the war. Column 2 shows estimates of posting pro-Russia content conditional on tweeting about the war using Equation 1. Column 3 reports estimates of posting any war-related content, again via Equation 1. We compute the approximate percentage change as the change from the pre-ban average. All specifications include user and day fixed effects, and standard errors are clustered at the user level.

FIGURE I.2
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 Equation 3 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 use the user-day panel dataset and include only user-day observations where the user was active in our sample as described in Section 3.2. We include only observations between February 22$^{\mathrm{nd}}$ and March 15$^{\mathrm{th}}$, 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 standard errors are clustered at the user level.

J Alternative Pro-Russia Source: TASS

In this Appendix, we explore a question that is complementary to our main analysis: 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 media environment by prompting users to seek out, or step into, alternative sources of pro-Russia content? The answer matters. If the demand for such content remained stable in the short term, as we expect it did, then the removal of key suppliers may have opened up a vacuum. And if the demand is inelastic in the face of top-down interventions, like a ban, that vacuum might quickly be filled by new or previously marginal actors. In what follows, 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 J.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 simple difference-in-difference analysis following Equation 2, with TASS designated as the treatment group. The coefficients should thus be interpreted as the differential impact of the ban on TASS, relative to its impact on Russia Today and Sputnik. 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 an engagement index. This index is constructed by standardizing the number of retweets, likes, and replies received by pro-Russia tweets. 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 3 and 5 include a trend capturing the level of activity, and columns 3 and 6 include day fixed-effects.

The results paint a clear picture: the ban did not affect TASS' activity or engagement levels relative to Russia Today and Sputnik. Somewhat unexpectedly, TASS did not adjust the volume of pro-Russia tweets following the ban, as none of the estimated models yield statistically significant differences. As for engagement, if anything, there appears to be a slight decline in interactions with TASS posts compared to those from Russia Today and Sputnik, but this effect diminishes once we control for activity trends, and it disappears entirely when day fixed effects are included. Overall, we find no evidence of institutional substitution. The void left by Russia Today and Sputnik was not filled by another state-backed source like TASS.

TABLE J.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	-0.340	-0.328	-0.314	0.255	0.273	0.171
Today and Sputnik	(0.223)	(0.223)	(0.222)	(0.206)	(0.204)	(0.238)
	[0.127]	[0.141]	[0.157]	[0.216]	[0.182]	[0.473]
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	-28.85	-27.95	-26.92	29.10	31.33	18.66
Observations	4321	4321	4321	796	796	796

Notes: The table examines the effect of the ban on TASS' activity and the engagement it received. It reports results from Poisson-Pseudo Maximum Likelihood regressions estimated in a two-way fixed effects difference-in-difference framework. The analysis is conducted at the tweet level, treating TASS as the exposed group and Russia Today and Sputnik as the control. The coefficients capture the differential impact of the ban on TASS relative to the banned outlets, which, as a reminder, remained active after the policy was introduced. 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 $e^{\beta} - 1$.

K Effect of the Ban on Demand for Russian Culture

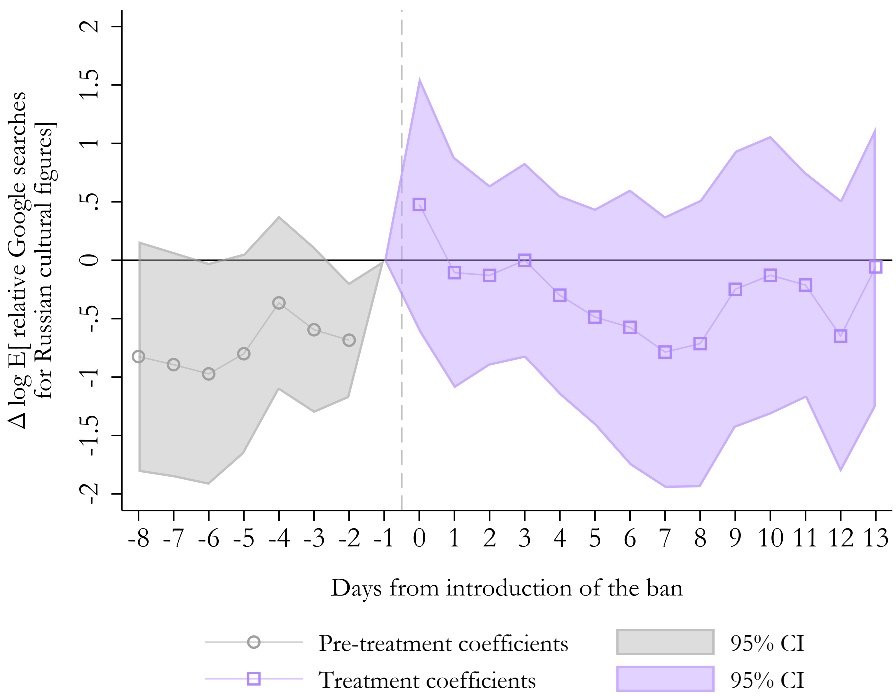
In this section, we study whether the ban led to any changes in the perception of Russia in the European countries that we study in our main analysis. To test this notion, we use the same difference-in-difference design as for the rest of the paper, but with the outcome of interest capturing demand for Russian culture. We operationalize demand for Russian culture via the relative amount of Google searches for nine representative Russian cultural figures. Specifically, we focus on the following cultural figures: Tolstoy, Dostoevsky, Tchaikovsky, Pushkin, Stravinsky, Nabokov, Kandinsky, Netrebko, and Sorokin. Importantly, we include both historical and contemporary artists.

For each of these figures, we take their relative Google search frequency in the UK, France, Germany, and Italy in the 21-day window around the ban used throughout the paper. In this exercise, we focus on the nine chosen figures and four countries due to issues with matching figures across all countries that return non-missing values for their relative Google searches. To extract relative Google searches, we use Google Trends data benchmarked via the `gtab` Python package (West 2020), which enables simplified comparisons of Google Trends queries.

One default option of Google Trends data is the comparison of Google Search queries within a single country over a certain time window. This allows for the comparison of the relative search frequency of up to five queries at once. With, for example, daily frequencies being of interest, the data is then organized such that the term which had the highest total searches of all term-day combinations is assigned the value 100. All other term-day combinations are assigned values relative to this maximum. This creates potential issues with respect to the sensitivity of the data and the included search terms. Additionally, Google Trends reports data rounded to the nearest integer, causing time series with outliers at the top to otherwise consist mainly of zeroes. To allow for better comparability and robustness, we set up a set of terms used for comparisons for each of our four countries of interest in the specified time window via `gtab`.

We report the event study results of this exercise in Figure K.1. We find no statistically significant impact of the ban on the demand for Russian cultural figures in EU Google searches relative to the demand in the UK. While the results of this simple exercise should not be overinterpreted, they suggest that the ban did not lead to a reduced interest in Russian cultural figures, at least as proxied by Google Search behavior.

FIGURE K.1
DAILY EVENT STUDY: IMPACT OF THE BAN ON GOOGLE SEARCHES FOR RUSSIAN CULTURAL FIGURES



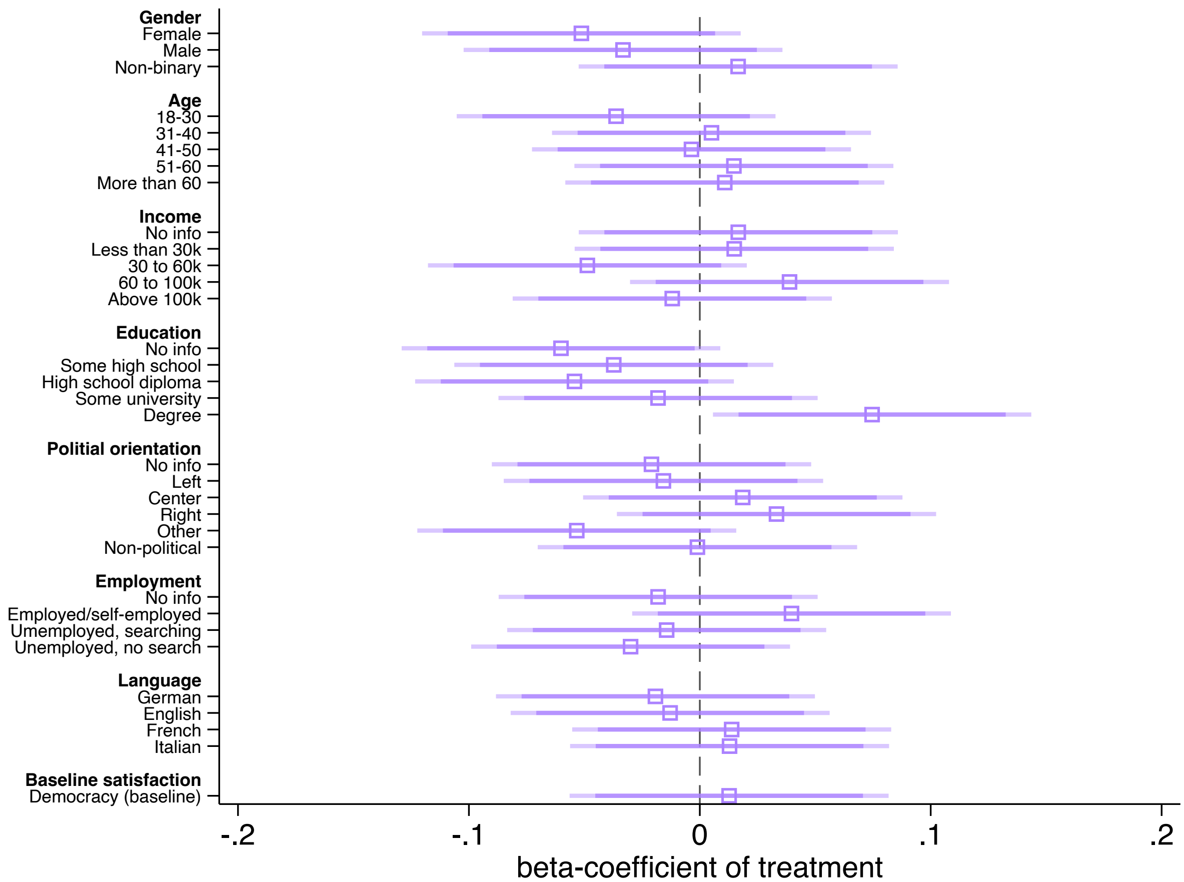
Notes: The figure displays coefficients and 95% confidence intervals from estimating Equation 4. The regression uses a daily panel dataset of relative Google searches for Russian cultural figures in France, Germany, Italy (treatment group) and the United Kingdom (control group). To mirror our other estimations, we only include observations between February 22nd and March 15th. As we report raw coefficients, the effects should be interpreted as changes in the logarithm of the expected value of the relative search volume (normalized between 0 and 100) for Russian cultural figures. The reported regression includes figure and day fixed effects, and standard errors are clustered at the figure 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.

L Experiment

In this Appendix, we provide additional evidence from our survey experiment reported in Section 6. Figure L.1 reports balance of sociodemographic characteristics with respect to receiving treatment. Figure L.2 reports heterogeneity of the treatment effect by baseline satisfaction with democracy in EU. Table L.1 reports results of the main outcomes sequentially including controls. Tables L.2 and L.3 report treatment effects on other, secondary outcomes and filler questions as a robustness check.

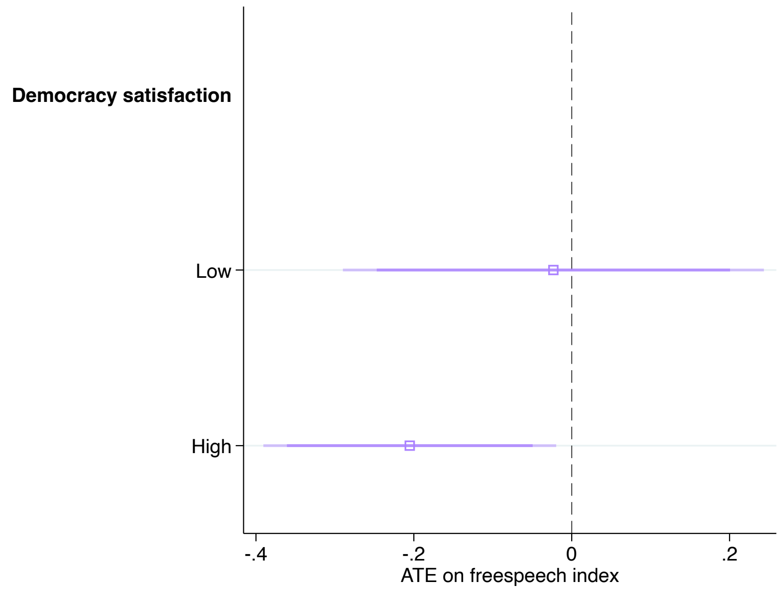
L.1 Additional Results

FIGURE L.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 6 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 L.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 6 for details) by baseline satisfaction with the functioning of democracy in the EU of the respondent. Dependent variable is the *Freespeech Index* computed as the average of approval to the *Freedom of speech* and *Media Independence* items. All items are using a 7-point Likert scale of approval of whether the European Union protects the corresponding construct. *Media Independence* was reverse coded. 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 L.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	-0.128	-0.115	-0.167	-0.057	-0.082
	(0.093)	(0.087)	(0.102)	(0.097)	(0.112)	(0.108)
	[0.333]	[0.141]	[0.258]	[0.086]	[0.613]	[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
Approx. Percentage Change	-8.60	-11.99	-12.38	-17.90	-182.75	-264.45
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 6 for details). Columns 1 and 2 use the pre-registered *Freespeech Index* as the dependent variable. It is computed as the average of approval to the *Freedom of speech* and *Media Independence* items recorded separately in columns 3-4 and 5-6, respectively. All items are using a 7-point Likert scale of approval of whether the European Union protects the corresponding construct. *Media Independence* was reverse coded. 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. We sequentially 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 L.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
Approx. Percentage Change	0.35	-1.85	4.71	3.96	-1.48	4.79
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 6 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. Columns 3 to 6 use items with 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 L.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
Approx. Percentage Change	-1.59	6.49	21.92	152.89
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 6 for details) on filler questions. Columns 1 and 4 use items with a 7-point Likert scale of approval of whether the European Union makes use of the corresponding measure in a sufficient manner as the dependent variable. 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.

L.2 Questionnaire

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)	o	o	o	o	o	o	o
The EU respects democratic norms even when under pressure. (2)	o	o	o	o	o	o	o
I trust the EU to uphold citizens' fundamental rights in times of crisis. (3)	o	o	o	o	o	o	o
The EU does not guarantee the independence of the media. (4)	o	o	o	o	o	o	o

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)	Assolutamer 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

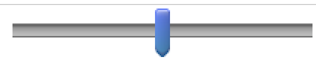
◯ Tendo a non fidarmi (1)

◯ 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)