

## Constructive Aggression? Multiple Roles of Aggressive Content in Political Discourse on Russian YouTube

Bodrunova, Svetlana S.; Litvinenko, Anna; Blekanov, Ivan; Nepiyushchikh, Dmitry

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

### Empfohlene Zitierung / Suggested Citation:

Bodrunova, S. S., Litvinenko, A., Blekanov, I., & Nepiyushchikh, D. (2021). Constructive Aggression? Multiple Roles of Aggressive Content in Political Discourse on Russian YouTube. *Media and Communication*, 9(1), 181-194. <https://doi.org/10.17645/mac.v9i1.3469>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

Article

## Constructive Aggression? Multiple Roles of Aggressive Content in Political Discourse on Russian YouTube

Svetlana S. Bodrunova<sup>1</sup>, Anna Litvinenko<sup>2,\*</sup>, Ivan Blekanov<sup>3,4</sup> and Dmitry Nepiyushchikh<sup>4</sup>

<sup>1</sup> School of Journalism and Mass Communications, St. Petersburg State University, 199004 St. Petersburg, Russia;  
E-Mail: s.bodrunova@spbu.ru

<sup>2</sup> Institute for Media and Communication Studies, Freie Universität Berlin, 14195 Berlin, Germany;  
E-Mail: anna.litvinenko@fu-berlin.de

<sup>3</sup> School of Mathematics and computer Science, Yan’an University, Yan’an Shanxi, China;

<sup>4</sup> Faculty of Applied Mathematics and Control Processes, St. Petersburg State University, 199004 St. Petersburg, Russia;  
E-Mails: i.blekanov@spbu.ru (I.B.), st075408@student.spbu.ru (D.N.)

\* Corresponding author

Submitted: 13 July 2020 | Accepted: 19 August 2020 | Published: 3 February 2021

### Abstract

Today, aggressive verbal behavior is generally perceived as a threat to integrity and democratic quality of public discussions, including those online. However, we argue that, in more restrictive political regimes, communicative aggression may play constructive roles in both discussion dynamics and empowerment of political groups. This might be especially true for restrictive political and legal environments like Russia, where obscene speech is prohibited by law in registered media and the political environment does not give much space for voicing discontent. Taking Russian YouTube as an example, we explore the roles of two under-researched types of communicative aggression—obscene speech and politically motivated hate speech—within the publics of video commenters. For that, we use the case of the Moscow protests of 2019 against non-admission of independent and oppositional candidates to run for the Moscow city parliament. The sample of over 77,000 comments for 13 videos of more than 100,000 views has undergone pre-processing and vocabulary-based detection of aggression. To assess the impact of hate speech upon the dynamics of the discussions, we have used Granger tests and assessment of discussion histograms; we have also assessed the selected groups of posts in an exploratory manner. Our findings demonstrate that communicative aggression helps to express immediate support and solidarity. It also contextualizes the criticism towards both the authorities and regime challengers, as well as demarcates the counter-public.

### Keywords

communicative aggression; hate speech; networked discussions; obscene speech; political protest; Russia; verbal aggression; YouTube

### Issue

This article is part of the issue “Dark Participation in Online Communication: The World of the Wicked Web” edited by Thorsten Quandt (University of Münster, Germany).

© 2021 by the authors; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

### 1. Introduction

In contemporary networked discussions, aggressive verbal behavior is a widespread phenomenon. The initial optimism about the democratic potential of online communicative milieus as new deliberative spaces (Diakopoulos & Naaman, 2011) was replaced in the

2010s by a pessimistic perception of them as dominated by trivia (Fuchs, 2017) as well as incivility, false information, and hate speech, conceptualized as ‘dark participation’ (Quandt, 2018). Aggressive speech has, for years, been almost exclusively considered a negative phenomenon worth detecting and filtering out, as aggression challenges the argumentative integrity of online

discussions (Vollhardt, Coutin, Staub, Weiss, & Deflander, 2007) and their normatively understood democratic quality (Cortese, 2006). In this capacity, it is simultaneously a digital threat of a non-political nature (Salter, Kuehn, Berentson-Shaw, & Elliott, 2019) and a threat to a rational and politically relevant public sphere (Miller & Vaccari, 2020; Pfetsch, 2018).

So far, this view upon aggressive speech has been challenged from several viewpoints. Thus, many works have addressed the dilemma of ‘free speech vs. hate speech’ (Hare & Weinstein, 2010; Howard, 2017; Weinstein, 2017). This well-known debate, including how hate speech and its being banned from public use relates to equality, autonomy, and legitimacy, has been reconstructed by Massaro (1990) and Waldron (2012). As a rule, bans on hate speech may be found in laws that prevent group hate and promote inter-group tolerance (Waldron, 2012, p. 8). However, this line of debate does not, in effect, challenge the understanding of aggressive speech as a threat, and only a threat, to democratic discussion; here, only the boundaries of what may be banned are debated. Also, it does not address the issue of politically motivated offensive language, as, in democracies, political groups are not considered disadvantaged minorities.

Only a few studies have so far tackled the issue of aggressive content as a form of individual/group empowerment or discussion fuel, while linguistic literature on offensive speech points to its positive functions for the speaker, such as release from tension or marking group belonging. Burns (2008, p. 61) has stated that “this type of linguistic behavior reflects and supports both the successful functioning of societies and individuals.” This might be especially true for restrictive political and legal environments where various types of public offense are prohibited by law and the political environment does not give much space for voicing discontent. In these contexts, dark participation could be a way to voice political dissent and rebel against the hegemonic discourses of the public sphere.

Our article aims at exploring the roles of aggressive language in political discussions in a so-far heavily under-researched context of countries with no sustainable democratic tradition. For such a study, today’s Russia represents a nearly perfect case. First, Russian society and Russian public communication of the 2010s have been fundamentally fragmented and increasingly polarized (Bodrunova & Litvinenko, 2015). Scholars describe a post-perestroika values-based division of the nation into a dominant traditionalist majority and an outlier minority of a mostly liberal-oppositional stance (Berezuev & Zvonareva, 2019), with a high level of mutual hostility. This allows for the exploration of the phenomenon of politically motivated hate speech. Second, the recent tightening of the political regime has, inter alia, brought along new bans on extremist speech, public swearing, and offending civil servants, thus giving offensive language the new connotation of an act of political disobedience. Third, in addition to hate lexicons, Russian has a

highly developed obscene sublanguage (*mat*), the functions of which go far beyond just expression of aggression, and which is even considered an “exceptionally rich” “linguistic system in its own right” and “a special genre of folk-art” (Dreizin & Priestly, 1982, p. 233–234).

Fourth, the Russian-speaking segment of Internet, or Runet, grew intensely and remained relatively unregulated for quite a long time before the mid-2010s (Vendil Pallin, 2017). By 2020, Internet penetration in Russia has reached 79% and was expected to exceed 100 million citizens by the end of 2020 (data by Krivoschapko, 2020), with 70.5% of people using mobile Internet (Elagina, 2020). On the one hand, since the mid-2000s, habits of discussions free from any bans have formed in Runet (Bodrunova, in press). To this day, Runet talk remains largely unbound by legal limitations, and the aforementioned restrictions introduced in the 2010s started to have their impact only very recently. For over two decades, Runet has served as a constellation of arenas in the online public sphere ‘parallel’ to the offline media landscape dominated by pro-state and loyal actors (Kiriyu, 2014), including arenas with alternative agendas (Bodrunova & Litvinenko, 2015) and large publics critical to leadership (Toepfl, 2020). On the other hand, Runet is known for its platform-wide echo chambering where, for example, Facebook is recognized as a liberal-oppositional filter bubble (Bodrunova & Litvinenko, 2015), while, in some Twitter discussions, nationalist discourses are dominant (Bodrunova, Blekanov, Smoliarova, & Litvinenko, 2019). Scholars also documented radicalization of Russian-language online speech quite early (Salimovsky & Ermakova, 2011).

Most studies of aggressive verbal behavior on Runet have focused on how to conceptualize, detect, and filter it out from ongoing or past online discussions (Koltsova, 2019). However, politicization of public swearing combined with wide de-tabooing of *mat* by cultural communities, the rural populace, and younger generations calls for reassessment of the possible roles of aggressive speech in political discussions. In particular, we ask what roles different types of aggressive language play in Runet political discussions.

For our enquiry, we use the case of Moscow protests against non-admission of oppositional and independent candidates to the elections of the Moscow city parliament of July to September 2019. We have analyzed 77,847 comments under the most viewed YouTube videos on this topic. We have assessed the roles of various types of aggressive content in the dynamics of the discussion by quantitative and qualitative instruments, including Granger tests and interpretation of discussion histograms.

The remainder of the article is structured as follows. In Section 2, we reconstruct the relevant theories and contextual knowledge on YouTube discussions in Russia. In Section 3, we pose the research questions and hypotheses. In Section 4, we describe data collection, the sample, and research steps. Section 5 provides the results and discussion, including the discovered roles

of aggressive content. In the final section, we place the results in the wider context of the dark participation and communicative aggression studies.

## 2. Aggressive Speech in Runet Discussions: Theory and Context

### 2.1. *Communicative Aggression in Networked Discussions: Only a Threat?*

Scholars, lawyers, and public institutions have produced a myriad of definitions for verbal aggression, hate speech, and other illegitimate ways of expression (Brown, 2017a, 2017b). Many legal definitions, though, are narrowed down by listing particular social groups vulnerable to verbal hatred or aggressive beliefs, like in case of anti-Semitism (see, for example, Council of Europe, 1997, p. 107). As mentioned above, they do not include groups defined by political views. However, the rise of authoritarianism and polarization in many countries demands an extension of the understanding of aggressive speech online.

In our study, we use a wide definition of aggressive speech by Cohen-Almagor (2011, p. 1): It is “bias-motivated, hostile, malicious speech aimed at a person or a group of people because of some of their actual or perceived innate characteristics,” which might include political groups. Here, we use the terms ‘[acts of] verbal aggression’ and ‘aggressive speech’ interchangeably, although we well realize that aggressive speech is a manifestation of aggression. In order to tailor this definition to the focus of our research, we use two other approaches that frame our understanding of aggressive speech.

First, for a formal definition of types of aggressive communication, we use the concept of ‘communicative aggression,’ as elaborated by Sidorov (2018). This umbrella concept allows for systematizing various forms of aggression in mediated communication. Following Sidorov’s logic, we argue that distinguishable pragmatic types of communicative aggression, such as cyberbullying, virtual racism, political hate speech, or swearing, link a certain speaker’s goal to a certain lexicon. Below, we operationalize the types of communicative aggression relevant for conflictual political discussions. Second, Parekh (2012), as cited and commented upon in Howard (2017), provides criteria for differentiating communicative aggression from political opinion, often also sharp and provocative. Parekh (2012, p. 41) has noted that hate speech “stigmatizes the target group by implicitly or explicitly ascribing to it qualities widely regarded as undesirable” and objectifies this group as a legitimate object of hostility. This approach allows us to select, within types of aggression, the lexical conglomerates linked to undesirable objects, events, or features. This approach also allows for distinction between verbal aggression and expressions of anger and other negative emotions. Undoubtedly, acts of verbal aggression are often used to express anger and hatred, but not always: For instance,

an obscene lexicon that substitutes normal speech in affective circumstances could express nearly any emotion, from disappointment to puzzlement to even joy. Our focus, thus, is not on emotions but on the speech used for humiliation and offence.

The effects of hate speech on political discussions have, so far, been almost exclusively assessed negatively (Van Aken, Risch, Krestel, & Löser, 2018). However, in certain contexts, aggressive content might help minorities voice political dissent not heard otherwise. Before the era of social networks, Delgado and Stefancic (1995) claimed that, among students, expressions of hatred may spur substantial on-campus debates on social discrimination. Davidson, Warmsley, Macy, and Weber (2017) have pointed to the possible use of potentially offensive language in a positive sense in groups that face discrimination, such as the LGBTQ community.

With the rise of social networking platforms, the issue of communicative aggression within them has become sharp to an extent unprecedented offline, where speech is not usually anonymized and detached from the speaker. The freedom of ‘dark participation’ has become part of a wider growth of dissonant public spheres (Pfetsch, 2018) where users neither seek consensus nor limit themselves by norms of public speech. The structure of such discussions has already been conceptualized as affective (Papacharissi, 2015; that is, hardly rational or reflexive, mostly highly emotional and quick to react) and ad hoc (Bruns & Burgess, 2011), which may lead to quick dissipation after the trigger event is over and does not allow the commenting evolve into a meaningful discussion. However, it is exactly this state of the online discourse that calls for rethinking of the roles of aggressive content within it. As stated above, it may allow users to shape a wide variety of thoughts and feelings expressed in rational discussions in another manner.

Aggressive content might also, presumably, influence discussion dynamics. Thus, Platonov and Svetlov (2020) have found that negative posts on the social network VK.com provoke a larger number of comments than neutral and positive ones. Our earlier works have shown that anger and aggression in tweets might be related to discussion intensity and pivotal turns in its topicality (Bodrunova et al., 2020; Smoliarova, Bodrunova, Yakunin, Blekanov, & Maksimov, 2019).

Given this, we might expect that aggressive speech may play multiple (positive and negative) roles in online discussions, for both its dynamics and content.

### 2.2. *Aggressive Speech on Russian YouTube and the Moscow Protests of 2019*

#### 2.2.1. Aggressive Language on Runet in the 2010s

Reassessment of aggression online might be especially relevant for political cultures like Russia’s. As stated above, relatively liberal regulation of online communication in the 2000s (Vendil Pallin, 2017) left room for both

free political discussion and its radicalization. In Russia, the rise of online radical, extremist, and nationalist speech became a scholarly concern much earlier than in most democratic countries (Etling et al., 2010; Kronhaus, 2012; Salimovsky & Ermakova, 2011). Widespread de-tabooing of obscene language contributed to this process. After the protests of 2011–2012 and the Ukrainian crisis, a range of legal restrictions were introduced (Litvinenko & Toepfl, 2019), and hundreds of legal cases against online posting opened (Gabdulhakov, 2020). Thus, in 2014, swear words were banned from use in media and the arts (Federal Law, 2013; enacted in 2014). In 2016, the so-called ‘Yarovaya law package’ provided formal grounds for recognition of online speech as extremist (Federal Law, 2016a, 2016b). In 2019, two other laws expanded the ban of ‘disrespectful’ statements about representatives of state power to include online space (Federal Law, 2019a, 2019b). These laws scaled up the range of instruments that could be used by the elites to curb dissent online (Litvinenko & Toepfl, 2019). Despite this, the online discourse has largely remained free of taboos.

In the 2010s, most English-language studies of aggressive speech on Runet were focused on inter-ethnic hostility (Bodrunova, Koltsova, Koltcov, & Nikolenko, 2017; Koltsova, 2019) and online stigmatization practices (Dudina, Judina, & Platonov, 2019). Several critical studies have linked aggressive speech to activity of pro-governmental trolls (Zvereva, 2020). In rare Russian-language studies of verbal aggression online, authors have mostly raised the issue of “degraded online talk” (Salimovsky & Ermakova, 2011, p. 74), without empirically testing its possible functions in online discussions. The linguo-pragmatic functions of obscene language have been explored by linguists (up to 27 functions; Havryliv, 2017; Zhel’vis, 1997). However, in none of these studies has communicative aggression been linked to discussion dynamics or freedom of expression in a restrictive media environment.

### 2.2.2. *Mat* in Russian Culture and in Online Speech

In this context, the Russian swearing lexicon deserves special attention: “Russia has an incredibly rich and versatile swear sublanguage, called *mat* [emphasis added], which is based on four key stems” (Pilkington, 2014; see also Pluzer-Sarno, 2000), with several more stems equally tabooed in public speech. Due to its inflective nature, Russian provides these stems diverse opportunities to perform the functions of nouns, verbs, adjectives, adverbs, and even interjections via the use of prefixes and suffixes. As *mat* has undergone the strictest tabooing within the Russian lexicon, we can expect that it bears the highest level of aggression and, if used, changes the discussion fabric. Similar to other languages, Russian *mat* had experienced de-tabooing long before social networks were in place. In today’s oral and online speech, swearing performs a variety of constructive pragmatic

functions, such as an increase of emotionality; release from psychological tension, demonstration of relaxedness and independence of the speaker; demonstration of disregard to restrictions; marking in-group belonging, etc. (Kosov, 2011, p. 37).

In today’s online speech, *mat* unites with politically motivated hate speech, another under-explored type of aggressive language. Their combination produces hybrid pejorative neologisms directed at both political camps: For example, towards the liberal-oppositional camp, *майданутый* (*maidanuty*, or, in English, ‘f\*\*\*ed in head by the Ukrainian Maidan revolution’) and *либераст* (*liberast*, that is, ‘liberal pederast’) are used, while *кремлядь* (*kremlyad*, or ‘a prostitute from Kremlin’) and *пропагандон* (*propagandon*, or ‘propagandistic condom’) are used against the authorities and loyalist media. These very neologisms are a sign of a politicization of the obscene lexicon and mark the importance of both obscene language and politically motivated hate speech for online political discussions.

### 2.2.3. YouTube as an Alternative to TV: A Crossroads of Polarized Opinions

We have chosen YouTube for our research because, in the recent years, YouTube has moved to the front of political communication in Russia. Since 2017, it has been the third most popular Runet website, according to Google Russia, with a monthly reach of 26% of Russians (Polyakova, 2017). For Russians, YouTube has gradually become an alternative to the state-dominated TV channels (Litvinenko, in press), with politics being one of the popular topics.

A recent study of political discourse on Russian YouTube during the 2018 presidential campaign has shown that the ‘Popular’ section, featuring top Russian YouTube videos, was dominated by oppositional actors (Litvinenko, in press). The leading oppositional channels, like Navalny’s Alexey Navalny (with 3,94 million subscribers) and Navalny LIVE (1,89 million subscribers); pro-liberal independent media/journalist channels, like VDud’ by Yury Dud’ (7,8 million), Alexey Pivovarov’s Redaktsia (1,4 million), and TV Rain (1,37 million); human-interest news vloggers like Roman Usachev (2,36 million); and critical political vloggers like Anatoly Shariy (2,38 million) and kamikazedead (an estimated 1,5 million of subscribers, as the actual figures are hidden) are comparable in viewership to the main state-affiliated federal channels, like NTV (9,56 million), Pervy kanal (4,95 million), and Rossiya 1 (3,62 million) and outperform many channels of national entertainment TV. Navalny’s anti-corruption investigation published on YouTube in 2017 ‘Don’t call him Dimon’ (Navalny, 2017) gathered more than 37,5 million views (as of January 2021) and triggered nationwide street protests (Gorbachev, 2017).

Another advantage of YouTube is that, unlike Facebook or VK.com, this platform has not yet been the major focus of legal action against political activists. Also,

Russian YouTube culture has been influenced by highly popular rap battles where obscene and offensive speech are core. As a result, Russian YouTube of the late 2010s has grown into a freely speaking opinion crossroads with a predominance of oppositional agenda and audience; this constituted a suitable case for our research.

For our analysis, we have taken the case of the Moscow protests of summer 2019, a vivid example of polarization of Russian public communication. Non-admission of oppositional and independent candidates on ballots for the Moscow city parliament elections triggered peaceful demonstrations in July 2019. They were suppressed by police, which led to an escalation of the conflict throughout August and September 2019. During those months, videos featuring the riots were widely circulated on YouTube and attracted a large number of comments, politically split and manifestly aggressive.

### 3. Research Questions

In this article, we ask whether various types of communicative aggression play active roles in both the dynamics and content of online political discussions. Here, in general, our approach is of an exploratory nature; we orient to the multiple roles of obscene language but explore whether other types of aggressive content may play similar roles. We follow the argument by Thelwall (2018), who has argued for the use of a mixed-method and exploration-oriented approach for studies of YouTube comments.

In the following research question, we ask whether various types of communicative aggression can spur discussions, what are the patterns of their appearance in the course of discussions, and whether users tend to radicalize individually and in discussion micro-clusters:

RQ1: Does communicative aggression affect discussion dynamics?

Moreover, in the next research question, we ask whether communicative aggression may be linked to democratic functions of the public sphere, including fostering both cross-group dialogue and counter-publicity. We also have in mind the above-mentioned functions of aggressive speech, where one can distinguish between psychological functions (like individual release of tension), social-psychological functions (like marking belonging), and political functions (like struggle for privilege or power):

RQ2: What roles do various types of communicative aggression play in political discussions online?

### 4. Method and Sampling

#### 4.1. Data Collection

For our analysis, we have chosen the comments under the most popular YouTube videos about the Moscow protests of 2019. To form a sample, we tried three strate-

gies: assessing the 'Popular' section, examining the popular political accounts, and searching by protest-related keywords. As a result, we decided to focus not on accounts but on individual videos, and have selected the 15 most popular videos in the search results for 'Moscow protests' that reached over 100,000 views uploaded to YouTube between July 27 (the start of the active phase of the street protests) and September 8 (the election day). Of those, two videos did not allow for comment upload: *Radio Liberty* disabled comments, and comments for the video by the oppositional news outlet *Current Time* were automatically blocked for download. Thus, we crawled and downloaded the comments for 13 videos from ten accounts, which resulted in 77,847 comments altogether, the number of comments per video ranging from 538 to more than 42,400. This collection allowed us to gather the most intense discussion fragments on the protests within Russian social media and trace its dynamics over time.

#### 4.2. Methods of Data Processing and Research Steps

##### 4.2.1. (Re)Conceptualizing Aggressive Speech for Conflictual Political Discussions

Following what was stated above, we first needed to conceptualize the types of aggression and the lexical conglomerates within them for automated analysis. A preliminary reading of over 3,000 posts in the dataset led us to rejecting the usual types of aggression, including cyberbullying, inter-ethnic hostility, and homophobia, as they were not the focus of the discussion. We have formulated four types of communicative aggression, all politically relevant and tabooed/prohibited by law: 1) Humiliation, including politically motivated hate speech and discriminative expressions—this type is partly directed at authorities and police and, thus, might fall under the law on offence of civil servants; 2) radical political claims (similar to 'Carthage must be destroyed'), including calls for aggressive action against individuals or groups—this is prohibited, as it may be considered extremism or a call to overthrow the existing political system; and 3) obscene language (*mat* and equally tabooed lexemes), which is tabooed and prohibited by the law against public swearing.

Then, we have further narrowed down our research via assessing the possibilities for automated analysis. We have found that, in over the 3,000 posts of the preliminary study, the number of radical calls for action amounted to only several dozen and were very diverse in lexical and grammatical terms; this was not enough even for deep learning and was only suitable for manual analysis. Thus, politically motivated hate speech and obscene language became our major focus.

##### 4.2.2. Selection of the Lexicon for Automated Analysis

Out of the preliminary analysis, the following lexical clusters were manually detected in obscene speech:

*mat* with case endings; *mat* with flexions (nouns, pronouns, verbs, adjectives, adverbs, etc.); lexemes tabooed equally to *mat*, with flexions; *mat* altered not to break the law (using \*, @, ', '\$, etc.); euphemisms of *mat*; and a lexicon of the lower body and defecation, which is less tabooed.

The following lexical clusters were manually detected in politically motivated hate speech inside anti-state discourse, pointing to: the police, in comparisons to garbage, lapdogs of the regime, cosmonauts (due to the uniforms), and the KGB; institutions, mostly the 'United Russia' party and the State Duma; the regime, by comparing its rule to that of the Nazis, to the Tsarist gendarmerie, and to organized crime (e.g., mafia); the state-affiliated media as 'purchasable' or bot-like; and individual politicians and their visual appearance.

The following lexical clusters were manually detected in politically motivated hate speech inside anti-opposition discourse: pejoratives of 'liberal' and 'democracy' made via flexions and stem combinations; 'fifth column,' the insider betrayers; diminutives and pejoratives formed out of the names of oppositional leaders Alexey Navalny and Luybov Sobol; linking the opposition to the USA, its grant funding, and philanthropists like George Soros; the Ukrainian pro-European Maidan revolution of 2013; LGBTQ and *gayropa* ('gay-favoring Europe'); and slacktivism.

We also detected offensive lexicon addressed to both political camps: markers of stupidity and ugliness, and prison slang; and to other groups: ethnic pejoratives.

Pre-tests on detection of aggression showed that the two latter groups were minor and did not play any significant role in the fabric of the discussion, while the three former groups mattered. Thus, the last two groups were excluded from further analysis.

The overall number of stems submitted to the stem detector amounted to 286. Also, 216 of them potentially contained flexions and case endings; thus, for the three types of aggressive speech, the overall expected number of lexical units was over 500.

#### 4.2.3. Addressing the RQs

To assess the role of aggression in discussion dynamics, we created the scripts to automatically detect the selected lexicons in the dataset and marked the comments as neutral, obscene, anti-state, or anti-opposition. The cases where the lexicon originally directed against the state or the opposition was used to form the opposite meaning (e.g., sarcastic) were fewer than 2% for all the videos; the comments where both anti-state and anti-opposition lexicons were found were even fewer, which shows that these lexicons are, indeed, markers of commenters' distinct political positioning.

Then, we conducted Granger tests juxtaposing the overall number of posts and the number of neutral posts with the number of the following types of posts: aggressive, obscene, politically aggressive, anti-state, and anti-

opposition, where 'politically aggressive' was anti-state and anti-opposition combined. As the discussion was intense for around four days in all the cases, three-day and four-day periods were selected for testing. We tested one-hour, two-hour, and six-hour increments; in addition, 15-minute and 30-minute increments were tested for the most commented-on videos. Altogether, 34 tests were performed. For the Granger tests, the comments that contained both anti-state and anti-opposition lexicons were excluded from analysis.

To assess individual vs. micro-group patterns of presence of aggressive speech, we decided to go for linear and interval histograms. While, for the Granger tests, the posts were singled out independently, for the histograms, we first singled out the posts with obscene language, then those with anti-state claims, and then those with anti-opposition claims; if the marking overlapped, the post was marked preferentially red, then yellow, then green. The number of overlapping posts, again, never grew over 3% of all aggressive posts. Hourly steps were introduced to demonstrate the discussion dynamics. Axis x ranged the users by time of their first post; axis y ranged the discussion in time. This allowed for assessing individual and micro-group user speech in time.

RQ2 was addressed in a grounded-theory manner. We did not pre-suppose any roles in advance, but were orienting to the lists of possible functions of obscene and aggressive speech mentioned above and our results from RQ1.

## 5. Results

### 5.1. RQ1

The videos that happened to be in our collection belonged to three types: foreign media in Russian (*BBC News*, *Euronews*, *DW*, and a Ukrainian correspondent in Moscow), Russian oppositional public political figures also active in media/online (Alexey Navalny, Maxim Shevchenko), and activist channels focused mainly on independent news production (*My Protest*, *Mordor Channel*, and *Superpower News*). As stated above, we have looked at whether aggressive posts fueled the subsequent discussion—or whether aggression just grew when the discussion itself grew, for which we used Granger tests. The results are presented in Figures 1 and 2 (two-hour increment). They show that, in 12 cases out of 13, the dynamics of the discussion was spurred by aggressive speech, and in 8 cases at least to the medium level. In at least five cases out of 13, this was not reciprocated—that is, if aggression grew or fell, the number of subsequent posts grew or fell accordingly. The videos showed different patterns of interaction between aggression and the overall discussion, but, in any case, we can make several conclusions.

Thus, both obscene and political aggression can play the role of discussion fuel; but, when one does it, the other is rarely involved. How aggression works seems

Videos	The overall number of posts in the discussion vs. communicative aggression									
	Direct effect					Reverse effect				
	All / aggressive	All / obscene	All / political	All / anti-state	All / anti-opposition	Aggressive / all	Obscene / all	Political / all	Anti-state / all	Anti-opposition / all
video_1	V STRONG	V STRONG	V STRONG	V STRONG	No	STRONG	No	MEDIUM	MEDIUM	No
video_2	No	No	No	No	No	MEDIUM	STRONG	STRONG	STRONG	V STRONG
video_3	No	WEAK	No	No	No	No	STRONG	No	No	No
video_4	No	No	No	WEAK	n/a	No	WEAK	No	No	n/a
video_5	No	No	No	No	n/a	No	WEAK	No	No	n/a
video_6	WEAK	No	No	No	No	V STRONG	MEDIUM	No	No	STRONG
video_7	WEAK	No	MEDIUM	MEDIUM	n/a	No	No	No	No	n/a
video_8	MEDIUM	MEDIUM	No	No	n/a	No	No	No	WEAK	n/a
video_9	No	No	No	STRONG	No	WEAK	WEAK	MEDIUM	WEAK	WEAK
video_10	No	No	No	No	No	MEDIUM	STRONG	STRONG	STRONG	STRONG
video_11	No	WEAK	No	n/a	n/a	No	No	STRONG	n/a	n/a
video_12	No	STRONG	No	No	n/a	No	No	WEAK	WEAK	n/a
video_13	MEDIUM	No	WEAK	n/a	n/a	V STRONG	No	V STRONG	n/a	n/a

Videos	The number of non-aggressive posts in the discussion vs. communicative aggression									
	Direct effect					Reverse effect				
	All / aggressive	All / obscene	All / political	All / anti-state	All / anti-opposition	Aggressive / all	Obscene / all	Political / all	Anti-state / all	Anti-opposition / all
video_1	V STRONG	V STRONG	V STRONG	V STRONG	No	STRONG	No	MEDIUM	MEDIUM	No
video_2	No	No	No	No	No	MEDIUM	STRONG	STRONG	STRONG	V STRONG
video_3	No	WEAK	No	No	No	No	STRONG	No	No	No
video_4	WEAK	No	No	WEAK	n/a	No	WEAK	No	No	n/a
video_5	No	No	WEAK	WEAK	n/a	No	WEAK	No	No	n/a
video_6	No	No	No	No	No	WEAK	MEDIUM	No	No	STRONG
video_7	WEAK	No	MEDIUM	STRONG	n/a	No	No	No	No	n/a
video_8	MEDIUM	MEDIUM	No	No	n/a	No	No	No	WEAK	n/a
video_9	WEAK	No	WEAK	STRONG	No	WEAK	WEAK	MEDIUM	WEAK	WEAK
video_10	No	No	No	No	No	MEDIUM	STRONG	STRONG	STRONG	STRONG
video_11	No	WEAK	No	n/a	n/a	No	No	No	n/a	n/a
video_12	No	MEDIUM	No	WEAK	n/a	No	No	WEAK	WEAK	n/a
video_13	WEAK	No	No	n/a	n/a	No	No	V STRONG	n/a	n/a

**Figure 1.** The results of the Granger test (two-hour increment). Notes: ‘Direct effect’ appears when the number of aggressive posts grows due to the overall growth of a discussion; ‘reverse effect’ appears when a discussion grows due to growth of the number of aggressive posts. 1) Blue: direct effect; orange: reverse effect; grey: data too scarce for the test. 2) Weak:  $F < 10$ ,  $0.01 < p \leq 0.05$ ; medium:  $p = 0.001 \leq p \leq 0.01$ ; strong:  $F > 10$ ,  $0.0001 \leq p < 0.001$ ; very strong:  $F > 10$ ,  $1e-12 < p < 0.0001$ .

to depend on two dimensions: author type and genre. Thus, aggression clearly fuels discussion under commentaries by political activists (even if the size of the discussion under Navalny’s video makes the trend reciprocal), while under commentaries by foreign news media both effects appear in weak to medium state, and, under news and the interview, the effects are clearly much less sound. For political celebrities, discussion fuels aggression (again, reciprocated under Navalny’s video)—that

is, the volume of the discussion matters. The smaller-sample cases show that obscene language might be provocative and bring on politically harsh speech, while, in bigger discussions, this effect is overcome. Due to scarcity, anti-liberal talk did not have an impact upon the discussions, but, overall, aggression played a significant role in how half of the discussions developed.

However, somewhat surprisingly, it is doubtful that one can call the comment conglomerates ‘discussions.’



#	Video Metadata								The overall number of posts					Non-aggressive posts				
	Video # in the dataset	Days assessed	Upload date	Overall Nposts	Channel type	Channel name	Content type	Topic	Aggressive	Obscene	Political	Anti-state	Anti-opposition	Aggressive	Obscene	Political	Anti-state	Anti-opposition
1	video_7	4	August 3	5037	Personal account of a public politician	Maksim Shevchenko	Comment	The Moscow protests. A leftist view										
2	video_8	4	July 28	592	Foreign media	Euronews in Russian	News	Moscow after the protests										
3	video_11	3	August 3	538	Foreign media: Ukrainian journalist	Roman Tsymbalyuk	Comment	Streaming on the Moscow protests										
4	video_12	3	August 10	929	Foreign media	DW in Russian	News	Reporting on protests and arrests										
5	video_1	4	July 30	42511	Personal account of a public politician	Navalny LIVE	Comment	The Moscow protests. What next?										
6	video_13	3	July 30	1099	Foreign media	BBC News Russian	Interview	Interview with victims hurt at the protests										
8	video_3	4	July 27	2028	Foreign media	BBC News Russian	News	Mass arrests at the Moscow protests										
7	video_6	4	August 5	7312	Activist oppositional / 'News&entertainment'	Moy protest / My protest	Comment	Commenting on police action										
9	video_2	4	August 31	8901	Activist oppositional / 'Independent news'	Kanal Mordor / Mordor Channel	Comment	Call for violent action against police										
10	video_10	4	July 28	2756	Activist oppositional / Road police watchdog	Dvizhenie / Traffic	Comment	Criticism on the Moscow authorities										
11	video_9	4	August 3	1078	Foreign media	BBC News Russian	Comment	The Moscow protests. What next?										
12	video_4	4	July 29	1455	Foreign media	BBC News Russian	Comment	Consequences of the Moscow protest										
13	video_5	4	August 21	4161	Activist oppositional / 'Independent news'	Novosti Sverhderzhavy / Superpower News	News	Putin compares the protests to jillets jauns										

**Figure 2.** The mixed effects of communicative aggression on the dynamics of the overall number of posts and the number of neutral posts (two-hour increment), accompanied by video metadata. Note: Blue—direct effect; green—reciprocal (both direct and reverse) effect; orange—reverse effect; grey—data too scarce for the test.

In the aforementioned histograms (Figures 3 and 4), we reproduced the commenting patterns by showing each user's posts in time, to see if they commented multiple times, and the intensity of the discussion by hourly intervals. The histograms clearly point to the affective and ad hoc nature of commenting, as, for all the discussions except Navalny's (Figure 3), repeated commenting is very rare (Figure 4). And, despite that, the fabric of the discussions is imbued with aggression, regardless of legal prohibitions; moreover, it rarely looks like an exchange of offensive comments, but rather like expressive 'shouts to the air' that condemn one of the political sides. We have also spotted micro-outbursts of aggression in smaller cases (Figure 4, circled); in big discussions, though, when individual users post multiple times (Figure 3, circled), they are likely to be aggressive, but when several users post simultaneously in a dynamic micro-cluster (Figure 3, squared), they, against expectations, abandon aggressive discourse. Thus, users mostly come to express to the author their aggressive support and solidarity in condemnation, and they rarely address their anger to fellow commenters; if engaged in further talking, they talk non-aggressively.

5.2. RQ2

Beyond immediate expression of support and solidarity, aggressive speech performs an array of functions

that tie together the affective nature of the discussions and the outer context. The most evident function is, by nature, sociological, as use of politicized obscene speech and pejoratives fosters political identities of the polarized camps sharply demarcating 'us' from 'them' (Van Dijk, 1993). This is already evident from the anti-state and anti-opposition lexical conglomerates described above, where an obscene lexicon intertwines with political pejoratives to the extent of neologisms. Additionally, a 'normal' group is put in open opposition to a group marked as undesirable, in agreement with Parekh (2012):

Cop mutts, they go against the people, shameful!!!

Every normal person disrespects rogues, thieves, and organized criminal groups!

Police are the best. One must nip the white-stripe coup d'état in the bud. Mercenary liberast beasts should be jailed!

Destructive to potential consensus between political antagonists, anti-state hate speech becomes a constructive means of counter-public consolidation. Interestingly, predominance of statements that favor oppositional discourse makes pro-state commenters, whose opinions otherwise dominate in the country, take a defensive tone:

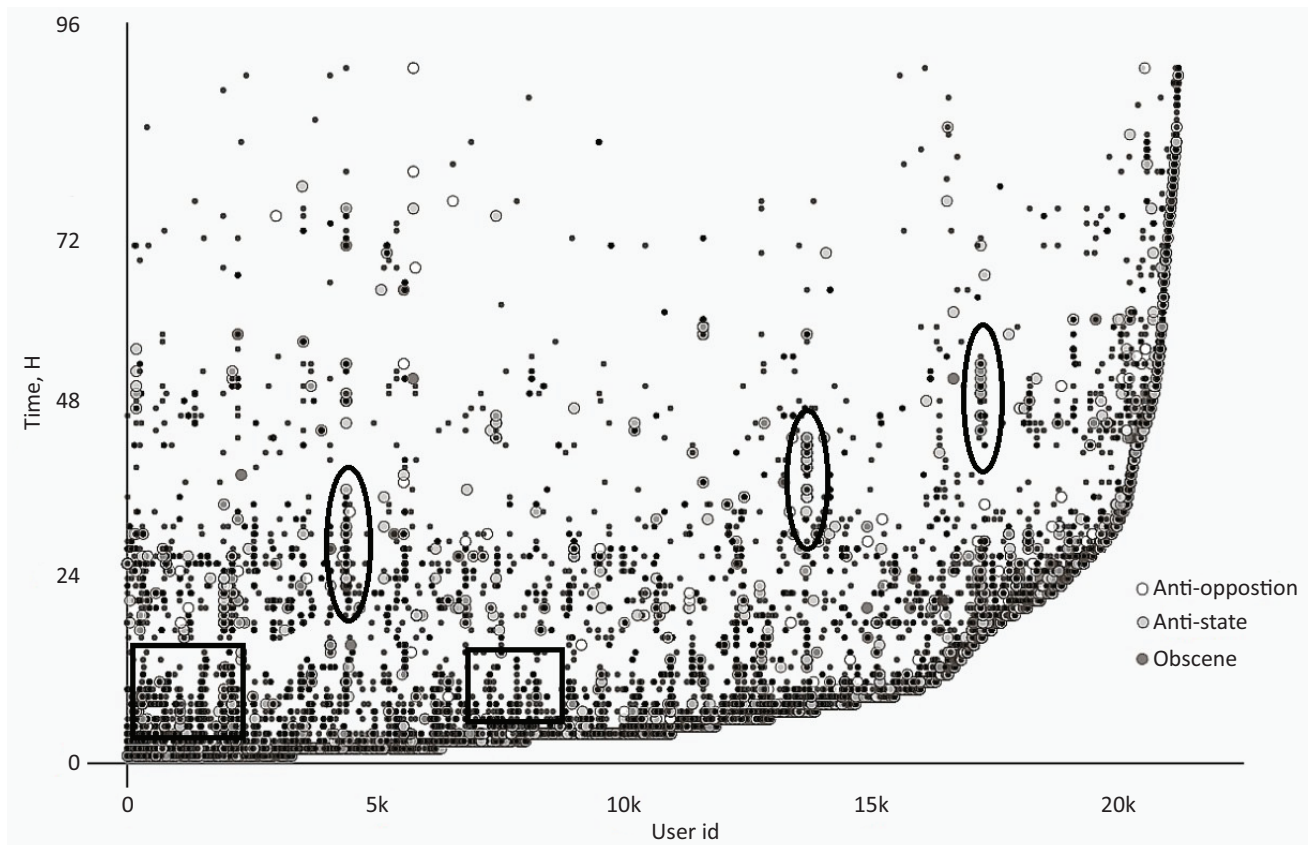


Figure 3. Interval histogram of the discussion, video 1: Individual aggressive vs. interactive neutral speech.

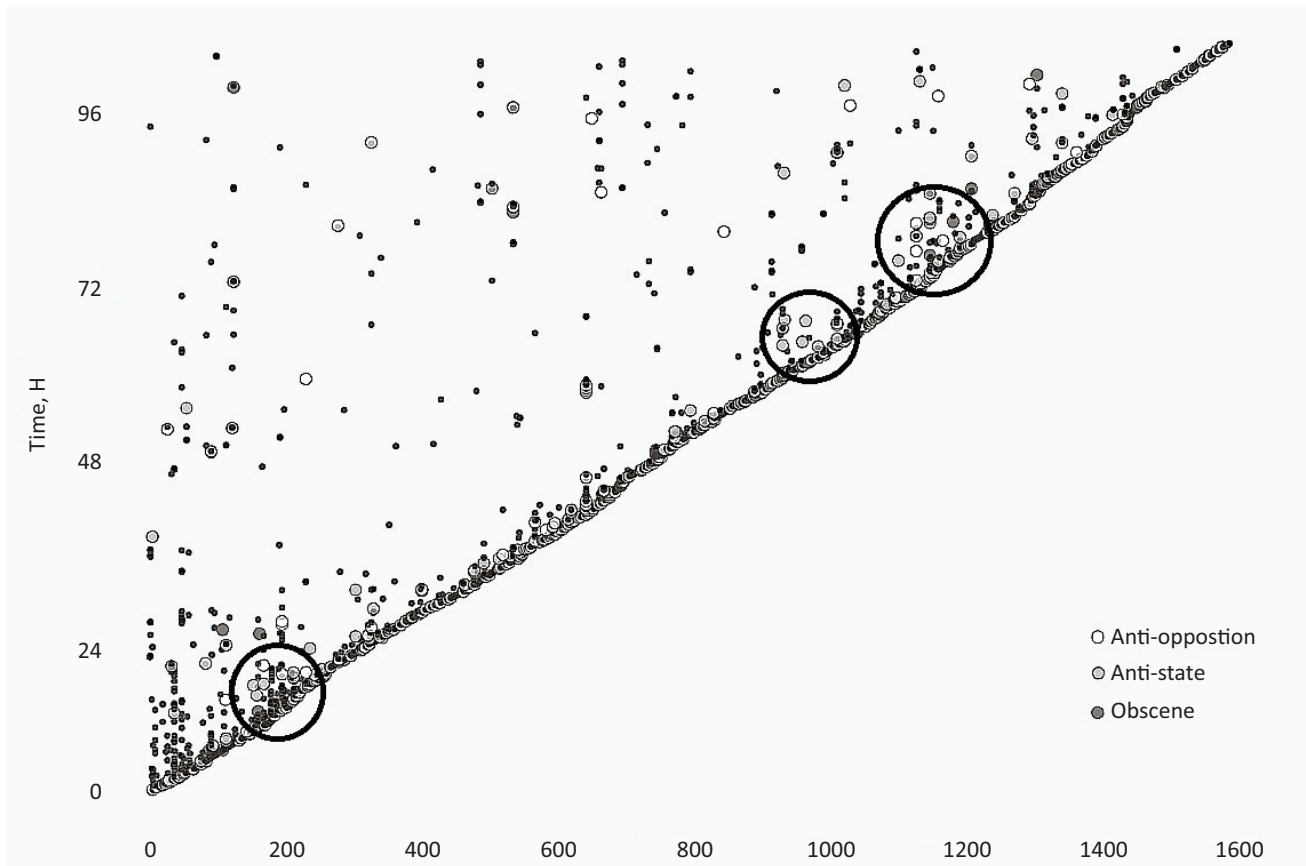


Figure 4. Linear histogram, video 13: Micro-clusters of aggression.

[To Lyubov Sobol whose hunger strike is depicted in the video] You think you put on glasses and got smart? Stupid. Realize one thing: Russia will not lie under your curators, even if you crap your pants together with the [US] State Dept.! As for your rallies...in the same Europe people like you are killed with fly swatters....Who do you serve, evil forces?

I am not going to prove anything. If you're idiot enough to be buying this whole show with Sobol's hunger strike, swollen Navalny, etc., it's up to you, but don't change flags after your idols leave for other countries.

In contrast, for the pro-opposition speakers, the very use of prohibited language may delineate a restriction-free realm. In political terms, aggressive speech challenges the hegemonic discourse (otherwise unchallengeable) and seeks to protect a way of thinking. And this, to an extent, is true for both political camps.

Via the use of particular metaphors in harsh lexicon, aggressive critique puts the current political regime in a row of national traumas from the 20th century, especially with World War II and the uncontrolled capitalism of the 'frantic 1990s,' united by the underlying reference to a dysfunctionality of the political scene captured by force:

Occupation, Siege, Nazism.

A 100% OCG [organized criminal group] runs the show. There are tops, and there are 'thugs.' Just the classic nineties but on a national scale!

Another metaphorical line points to the issue of freedom vs. excessive control:

In a concentration camp, there are no rights.

Democracy and freedom are the least of it. Russia has been captured by Nazis!!!

Bandits in uniform. They do not protect the citizens; they enslave them.

Anti-liberal speech, in its turn, radically addresses fears of yet another system revolt: post-Cold War cleavages in values, deep suspicion towards American and European intentions, and post-imperial resentment. Here, post-Maidan Ukraine is often posed as an example of a country destroyed by a 'color revolution,' and, likewise, the 1990s are a symbol of poverty and instability:

Go on rallying! [The USA] State Dept. will be grateful...1991 and 1993 forgotten? Moscow b\*\*\*\*es! Trying to sell Russia again? That time [it was] for jeans and salami, and now what for?

For a second Maidan! For the return to the 1990s! For hunger and destitution! Hurraaaaaa!!!!

While obscene language marks the emotion intensity, it is exactly the aggressive political talk that is responsible for contextualization—and, thus, for re-appropriation of the right of political interpretation for both camps. Thus, two types of aggression (obscene and political) support each other in user posts. Moreover, in the absence of sound public debate and chances for peaceful dissent, aggressive ‘speaking out what is suppressed’ serves as a way to vent anger. This, *inter alia*, can be beneficial for the regime, as users who vent their anger online may refrain from offline protest, as described by Toepfl (2020) in his conceptualization of risks and benefits of critical publics for autocracies.

Last but not least, we observed a high level of creativity in political use of tabooed language, from stem fusion to ‘bleeping’ of prohibited words. An opposition to the sterile and deliberate official language, it marks the grassroots, ingenuous, and censorship-free character of political discussions online.

## 6. Discussion and Conclusion

In this article, we have challenged the view of communicative aggression online as a necessarily negative means of disempowerment and de-legitimation. We have shown that under-studied types of communicative aggression, such as obscene language and politically motivated hate speech, may not only escalate conflict in public debate but, by contrast, also play constructive democratic roles for individuals, groups, and discursive processes in a restrictive political atmosphere. This includes expression of support and solidarity, counter-public consolidation, re-appropriation of rights to interpret historical context, and manifestation of creativity and wit that opposes restrictions and official discourses. Some of these functions are, of course, double-edged swords: Thus, consolidation of a counter-public leads to widening gaps between it and the wider public, and interpretation of context available for both pro- and counter-establishment groups may potentially lead to abuse of historical memory. However, such elements of public deliberation are in any case more characteristic of democratic contexts than of autocratic ones. This is why we claim that one can speak of ‘constructive aggression’ and differentiate it from destructive hate speech and an obscene lexicon if it plays constructive roles in public discussions; similarly, misinformation may be shared online with constructive motivations (Metzger, Flanagin, Mena, Jiang, & Wilson, 2021).

Our results have supported the claim that communicative aggression not only thrives within YouTube reactions to oppositional videos but also fuels user talk. We have spotted micro-outbursts of both aggression (by newcomers) and neutral talk (by discussants), while, in general, aggression was spread throughout the affective conglomerates of posts.

By analyzing the roles of aggressive content, we have seen that it may play a significant role in the formation of ad hoc counter-publics around a politically polarizing issue in online discussions, even when the patterns of user interaction reveal a disrupted public. It is politically motivated hate speech that, at least partly, forms the fabric of collective expression in comments. We have explored how the personal-level functions of aggressive speech, including *mat*, manifest themselves on the group level and gain political relevance. In the Russian case, communicative aggression is linked to giving voice to political opposition, which is otherwise excluded from the mainstream discourse, and may foster counter-publics and offline action, as was the case during the Russian protests of 2011–2012 (Bodrunova & Litvinenko, 2015). This function of aggressive speech is in line with the strand of research on agonistic public spheres that emphasizes the importance of political conflict and political voices ‘from the margins’ for public deliberation (Dahlberg, 2007, p. 128). If, as in Russia, offensive language is officially banned in the media, using this kind of language *per se* might become a way to challenge the hegemonic official discourses.

The limitations of our research come from its exploratory nature, as well as from the limited number of videos around which the core discussion evolved. But even more, they stem from the nature of the discovered publics, as the lack of interaction between commenters prevents the use of, for example, social network analytics. Our findings support the idea of affective and dissonant publics, but partly re-interpret dissonance and ‘dark participation’ as democratically functional. They also point out to cumulative effects in online communication.

Our research shows that fighting aggressive speech on global social media platforms can, *inter alia*, give autocrats a tool to curb political dissent online. It can also mean depriving marginalized groups of the opportunity to vent their anger, which may lead, instead of to the expected harmonization of political communication, to an escalation of violence, both online and offline. One may ask how true this might be for democratic contexts; this would be intriguing to explore in future research.

## Acknowledgments

This research has been supported in full by Russian Science Foundation, grant 16-18-10125-P.

## Conflict of Interests

The authors declare no conflict of interests.

## References

- Berezuev, E. A., & Zvonareva, L. V. (2019). Quasi-modernization in today’s Russia as a confrontation of values of modernism and traditionalism. *Manuscript*, 12(1), 93–98.

- Bodrunova, S. S. (in press). Information disorder practices in/by contemporary Russia. In H. Tumber & S. Waisbord (Eds.), *The Routledge companion to media misinformation and populism* (pp. 279–289). London: Routledge.
- Bodrunova, S. S., Blekanov, I., Smoliarova, A., & Litvinenko, A. (2019). Beyond left and right: Real-world political polarization in Twitter discussions on inter-ethnic conflicts. *Media and Communication*, 7(3), 119–132.
- Bodrunova, S. S., Koltsova, O., Koltcov, S., & Nikolenko, S. (2017). Who's bad? Attitudes toward resettlers from the post-Soviet south versus other nations in the Russian blogosphere. *International Journal of Communication*, 11, 3242–3264.
- Bodrunova, S. S., & Litvinenko, A. A. (2015). Four Russias in communication: Fragmentation of the Russian public sphere in the 2010s. In B. Dobek-Ostrowska & M. Glowacki (Eds.), *Democracy and media in Central and Eastern Europe* (pp. 63–79). Bern: Peter Lang.
- Bodrunova, S. S., Nigmatullina, K., Blekanov, I. S., Smoliarova, A., Zhuravleva, N., & Danilova, Y. (2020). When emotions grow: Cross-cultural differences in the role of emotions in the dynamics of conflictual discussions on social media. In G. Meiselwitz (Ed.), *Proceedings of the 12th international conference SCSM 2020* (Vol. 12194, pp. 433–441). Cham: Springer.
- Brown, A. (2017a). What is hate speech? Part 1: The myth of hate. *Law and Philosophy*, 36(4), 419–468.
- Brown, A. (2017b). What is hate speech? Part 2: Family resemblances. *Law and Philosophy*, 36(5), 561–613.
- Bruns, A., & Burgess, J. E. (2011). *The use of Twitter hashtags in the formation of ad hoc publics*. Paper presented at the European Consortium for Political Research (ECPR) General Conference 2011, Reykjavik, Iceland.
- Burns, M. C. (2008). Why we swear: The functions of offensive language. *Monash University Linguistics Papers*, 6(1), 61–69.
- Cohen-Almagor, R. (2011). Fighting hate and bigotry on the Internet. *Policy & Internet*, 3(3), 1–26.
- Cortese, A. J. P. (2006). *Opposing hate speech*. Westport, CT: Greenwood Publishing Group.
- Council of Europe. (1997). *Recommendation of The Committee of Ministers to member states on 'hate speech'* (No. R (97) 20). London: Council of Europe. Retrieved from <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680505d5b>
- Dahlberg, L. (2007). The Internet and discursive exclusion: From deliberative to agonistic public sphere theory. In L. Dahlberg & E. Siapera (Eds.), *Radical democracy and the Internet* (pp. 128–147). London: Palgrave Macmillan.
- Davidson, T., Warmesley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. In O. Varol, E. Ferrara, C. A. Davis, F. Menczer, & A. Flammini (Eds.), *Proceedings of the eleventh international AAAI conference on web and social media (ICWSM 2017)* (pp. 512–515). New York, NY: AAAI.
- Delgado, R., & Stefancic, J. (1995). Ten arguments against hate-speech regulation: How valid. *Law Review*, 23. <https://heinonline.org/HOL/LandingPage?handle=hein.journals/nkenlr23&div=32&id=&page=>
- Diakopoulos, N., & Naaman, M. (2011). Towards quality discourse in online news comments. In P. Hinds, J. Tang, J. E. Bardram, J. Wang, & N. Ducheneaut (Eds.), *Proceedings of the ACM 2011 conference on computer supported cooperative work* (pp. 133–142). New York, NY: ACM.
- Dreizin, F., & Priestly, T. (1982). A systematic approach to Russian obscene language. *Russian Linguistics*, 6, 233–249.
- Dudina, V., Judina, D., & Platonov, K. (2019). Personal illness experience in Russian social media: Between willingness to share and stigmatization. In S. El Yacoubi, F. Bagnoli, & G. Pacini (Eds.), *Proceedings of the 6th international conference on Internet science (INSCI 2019)* (Vol. 11938, pp. 47–58). Cham: Springer.
- Elagina, D. (2020, July 20). Internet penetration rate in Russia from September 2019 to February 2020, by type of device. *Statista*. Retrieved from <https://www.statista.com/statistics/1051507/russia-internet-penetration-by-medium/#statistic>
- Etling, B., Alexanyan, K., Kelly, J., Faris, R., Palfrey, J. G., & Gasser, U. (2010). *Public discourse in the Russian blogosphere: Mapping RuNet politics and mobilization* (Research Publication 2010–2011). Cambridge, MA: Harvard Berkman Center. Retrieved from [https://dash.harvard.edu/bitstream/handle/1/8789613/Public\\_Discourse\\_in\\_the\\_Russian\\_Blogosphere\\_2010.pdf?sequence=1&sa=U&ei=qGVQU8\\_NAoStyAS57oFg&ved=0CEAQFjAG&usq=AFQjCNEMhqq9I61iT2bGDO2ITuGgV\\_hOTw](https://dash.harvard.edu/bitstream/handle/1/8789613/Public_Discourse_in_the_Russian_Blogosphere_2010.pdf?sequence=1&sa=U&ei=qGVQU8_NAoStyAS57oFg&ved=0CEAQFjAG&usq=AFQjCNEMhqq9I61iT2bGDO2ITuGgV_hOTw)
- Federal Law. (2013). *On Amendments to Clause 4 of the Law of the Russian Federation 'On Mass Media' and Clause 13.21 of the Code of Administrative Offences of the Russian Federation: Federal law #34-FZ of April 5, 2013*. Moscow: Federal Law.
- Federal Law. (2016a). *On Amendments to the Federal Law 'On Countering Terrorism' and Certain Legislative Acts of the Russian Federation regarding the establishment of additional measures to counter terrorism and ensure public safety: Federal law #374-FZ of July 6, 2016*. Moscow: Federal Law.
- Federal Law. (2016b). *On Amendments to the Criminal Code of the Russian Federation and the Code of Criminal Procedure of the Russian Federation regarding the establishment of additional measures to counter terrorism and ensure public safety: Federal law #375-FZ of July 6, 2016*. Moscow: Federal Law.
- Federal Law. (2019a). *On Amendments to the Code of Administrative Offences of the Russian Federation:*

- Federal law #28-FZ of March 18, 2019. Moscow: Federal Law.
- Federal Law. (2019b). *On Amendments to the Federal Law 'On Information, Information Technologies and Information Protection': Federal law #30-FZ of March 18, 2019*. Moscow: Federal Law.
- Fuchs, C. (2017). *Social media: A critical introduction*. London: Sage.
- Gabdulhakov, R. (2020). (Con)trolling the Web: Social media user arrests, state-supported vigilantism and citizen counter-forces in Russia. *Global Crime*. <https://doi.org/10.1080/17440572.2020.1719836>
- Gorbachev, A. (2017, March 24). Pochemy na akzii protesta vyshlo stolko podrostkov? [Why have so many teenagers attended the protests?]. *Meduza*. Retrieved from <https://meduza.io/feature/2017/03/27/pochemu-na-aktsii-protesta-vyshlo-stolko-podrostkov-i-chego-oni-hotyat>
- Hare, I., & Weinstein, J. (Eds.). (2010). *Extreme speech and democracy*. Oxford: Oxford University Press.
- Havryliv, O. (2017). Verbale Aggression: das Spektrum der Funktionen [Verbal aggression: The spectrum of functions]. *Linguistik*, 82, 27–47.
- Howard, J. W. (2017). Free speech and hate speech. *Annual Review of Political Science*, 22, 93–109.
- Kiriya, I. (2014). Social media as a tool of political isolation in the Russian public sphere. *Journal of Print and Media Technology Research*, 3(2), 131–138.
- Koltsova, O. (2019). Methodological challenges for detecting interethnic hostility on social media. In S. S. Bodrunova, O. Koltsova, A. Følstad, H. Halpin, P. Kolozaridi, L. Yuldashev, . . . H. Niedermayer (Eds.), *Proceedings of international the 5<sup>th</sup> international conference on Internet science (INSCI 2018)* (Vol. 11551, pp. 7–18). Cham: Springer.
- Kosov, A. V. (2011). Mythological conscience and invective vocab. *Novosibirsk State University Bulletin*, 5(2), 35–40.
- Krivoshapko, Y. (2020, April 15). Runet audience is to break through to 100 million users in 2020. *Rossiyskaya Gazeta*. Retrieved from <https://rg.ru/2020/04/15/auditorii-a-runeta-v-2020-godu-probet-planku-v-100-mln-polzovatelej.html>
- Kronhaus, M. (2012). *The Russian language on the verge of a nervous breakdown*. Moscow: Astrel.
- Litvinenko, A. A. (in press). YouTube as alternative television in Russia: Political videos during the presidential election campaign 2018. *Social Media + Society*.
- Litvinenko, A. A., & Toepfl, F. (2019). The 'gardening' of an authoritarian public at large: How Russia's ruling elites transformed the country's media landscape after the 2011/12 protests 'For fair elections.' *Publizistik*, 64(2), 225–240.
- Massaro, T. M. (1990). Equality and freedom of expression: The hate speech dilemma. *William & Mary Law Review*, 32. <https://scholarship.law.wm.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1923&context=wmlr>
- Metzger, M. J., Flanagin, A. J., Mena, P., Jiang, S., & Wilson, C. (2021). From dark to light: The many shades of sharing misinformation online. *Media and Communication*, 9(1), 134–143.
- Miller, M. L., & Vaccari, C. (2020). Digital threats to democracy: Comparative lessons and possible remedies. *The International Journal of Press/Politics*. <https://doi.org/10.1177/1940161220922323>
- Navalny, A. (2017, March 2). *Don't call him Dimon* [Video file]. Retrieved from [https://www.youtube.com/watch?v=qrwk7\\_GF9g&ab\\_channel=%D0%90%D0%BB%D0%B5%D0%BA%D1%81%D0%B5%D0%B9%D0%9D%D0%B0%D0%B2%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9](https://www.youtube.com/watch?v=qrwk7_GF9g&ab_channel=%D0%90%D0%BB%D0%B5%D0%BA%D1%81%D0%B5%D0%B9%D0%9D%D0%B0%D0%B2%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9)
- Papacharissi, Z. (2015). *Affective publics: Sentiment, technology, and politics*. Oxford: Oxford University Press.
- Parekh, B. (2012). Is there a case for banning hate speech? In M. Herz & P. Molnar (Eds.), *The content and context of hate speech: Rethinking regulation and responses* (pp. 37–56). Cambridge: Cambridge University Press.
- Pfetsch, B. (2018). Dissonant and disconnected public spheres as challenge for political communication research. *Javnost-The Public*, 25(1/2), 59–65.
- Pilkington, A. (2014, May 8). The rich, swearsy sub-language that will protect Russia from Putin's latest crackdown. *The Conversation*. Retrieved from <https://theconversation.com/the-rich-sweary-sub-language-that-will-protect-russia-from-putins-latest-crackdown-26362>
- Platonov, K., & Svetlov, K. (2020). *Involvement in discussions on Khachaturian sisters' case on SNS Vkontakte: Agenda polarization online*. Paper presented at Networks in the Global World (NetGloW'2020) conference, St. Petersburg, Russia.
- Pluzer-Sarno, A. (2000). *Mat vocabulary as a phenomenon of Russian culture*. Moscow: Novaya russkaya kniga.
- Polyakova, V. (2017, December 6). YouTube zanyal tretje mesto po ohvatu auditorii v RuNete [YouTube takes third place in audience outreach on RuNet]. *Seonews.Ru*. Retrieved from <https://www.seonews.ru/events/YouTube-zanyal-trete-mesto-pokhvatu-auditorii-v-runete>
- Quandt, T. (2018). Dark participation. *Media and Communication*, 6(4), 36–48.
- Salimovsky, V. A., & Ermakova, L. M. (2011). Extremist discourse in Runet mass communication. *Rossiyskaya i zarubezhnaya filologiya*, 3(15), 71–80.
- Salter, L. A., Kuehn, K., Berentson-Shaw, J., & Elliott, M. (2019). Digital threats to democracy: Literature review part 1: Threats and opportunities. *Luminate Workshop*. Retrieved from <https://mro.massey.ac.nz/bitstream/handle/10179/14949/3.DD-background-paper-lit-review-1-WEB.pdf?sequence=1>
- Sidorov, V. A. (2018). Communicative aggressions of the 21st century: Definition and analysis of the prerequisites. *Vestnik*, 15(2), 300–311.

- Smoliarova, A. S., Bodrunova, S. S., Yakunin, A. V., Blekanov, I., & Maksimov, A. (2019). Detecting pivotal points in social conflicts via topic modeling of Twitter content. In S. S. Bodrunova, O. Koltsova, A. Følstad, H. Halpin, P. Kolozaridi, L. Yuldashev, . . . H. Niedermayer (Eds.), *Proceedings of the 5<sup>th</sup> international conference on Internet science (INSCI 2018)* (Vol. 11551, pp. 61–71). Cham: Springer.
- Thelwall, M. (2018). Social media analytics for YouTube comments: Potential and limitations. *International Journal of Social Research Methodology*, 21(3), 303–316.
- Toepfl, F. (2020). Comparing authoritarian publics: The benefits and risks of three types of publics for autocrats. *Communication Theory*, 30(2), 105–125.
- Van Aken, B., Risch, J., Krestel, R., & Löser, A. (2018). *Challenges for toxic comment classification: An in-depth error analysis*. Unpublished manuscript.
- Van Dijk, T. A. (1993). *Elite discourse and racism*. Newbury Park, CA: Sage Publications.
- Vendil Pallin, C. (2017). Internet control through ownership: The case of Russia. *Post-Soviet Affairs*, 33(1), 16–33.
- Vollhardt, J., Coutin, M., Staub, E., Weiss, G., & Deflander, J. (2007). Deconstructing hate speech in the DRC: A psychological media sensitization campaign. *Journal of Hate Studies*, 5(15), 15–35.
- Waldron, J. (2012). *The harm in hate speech*. Cambridge, MA: Harvard University Press.
- Weinstein, J. (2017). Hate speech bans, democracy, and political legitimacy. *Constitutional Commentary*, 32, 527–583.
- Zhel'vis, V. (1997). *Pole brani: Skvernoslovie kak sotsial'naya problema v yazykah i kul'turah mira* [Field of word battle: Swearing as a social problem in world languages and cultures]. Bryansk: Lodomir.
- Zvereva, V. (2020). 11 State propaganda and popular culture in the Russian-speaking Internet. In M. Wijermars & K. Lehtisaari (Eds.), *Freedom of expression in Russia's new mediasphere*. London: Routledge.

### About the Authors



**Svetlana S. Bodrunova** (Dr. Hab.) is Professor at School of Journalism and Mass Communications, St. Petersburg State University, Russia. She has led six research projects and published two books and over 80 research papers in Russian and English on Russian and European journalism, media and politics, social media, and ethnicity in communication. She leads the Center for International Media Research in her university.



**Anna Litvinenko** (PhD) is Researcher at Digitalization and Participation department of Institute for Media and Communication Studies, Freie Universität Berlin, Germany. After receiving her PhD in 2007, she worked as Associate Professor at the Department of International Journalism of St. Petersburg State University, Russia. Her research focusses on the interrelation of media and politics in the digital age and on the role of social media platforms in various socio-political contexts.



**Ivan Blekanov** has PhD in System Analysis, Control and Data Processing. He is Assistant Professor and Head of Department of Programming Technology at St. Petersburg State University, Russia. He has published over 40 academic publications in Russian and English and has lead data science teams within 5 interdisciplinary projects. His research interests include information retrieval and data mining, webometrics and web analytics, social network analysis and computer forensics. He also works as IT industry consultant.



**Dmitry Nepiyushchikh** is a Graduate Student at St. Petersburg State University, Russia. He has been involved into research projects that dealt with social media data collection and social network analysis. His research interests include crawling of web and social media, user text classification, and visualization of the results of data processing.