RUDN Journal of Political Science. ISSN 2313-1438 (print), ISSN 2313-1446 (online) 2022 Vol. 24 No. 3 480–498
Вестник Российского университета дружбы народов. Серия: ПОЛИТОЛОГИЯ http://journals.rudn.ru/political-science

ПОЛИТИЧЕСКАЯ ПОЛЯРИЗАЦИЯ И ИНТЕРНЕТ-ПРОТЕСТ

POLITICAL POLARIZATION AND INTERNET PROTEST

DOI: 10.22363/2313-1438-2022-24-3-480-498

Научная статья / Research article

Аффективная политическая поляризация и язык ненависти: созданы друг для друга?

Д.К. Стукал[1], А.С. Ахременко[1], А.П. Петров[2]

[1] Национальный исследовательский университет «Высшая школа экономики», Москва, Российская Федерация

[2] Институт прикладной математики им. М.В. Келдыша Российской академии наук, Москва, Российская Федерация

dstukal@hse.ru

Аннотация. Многочисленные исследования указывают на прогрессирующий рост показателей политической поляризации в странах мира в целом, а также ее разновидности — аффективной поляризации. Предшествующие работы, посвященные данной проблеме, опирались почти исключительно на реактивные методы исследования (включая опросы и экспериментальные методики), не позволяющие наблюдать за поведением объектов анализа в естественной среде. В данном исследовании мы предлагаем альтернативный подход, основанный на анализе наблюдаемого поведения пользователей социальных сетей и выявлении ключевых поляризующих расколов путем анализа использования языка вражды в отношении различных целевых групп. Предложена оригинальная методика кодировки текстовых сообщений, включающая два ключевых компонента: операционализированное определение языка вражды как явления, содержащего хотя бы один из трех признаков: оскорбление, дискриминация, агрессия; а также оригинальный гайд для кодирования случаев использования языка вражды. Предлагаемая методика апробируется в работе на эмпирическом материале, включающем более 5000 сообщений, опубликованных в социальной сети ВКонтакте по тематике встреч Президентов России и Беларуси в 2020–2021 гг. Была проведена кодировка собранных данных, и на ее основе проведен анализ, выявивший две устойчивые линии раскола, связанные с внутриполитическими размежеваниями в этих странах и противопоставлением России/Беларуси странам Запада, а также третью, соответствующую противопостав-

лению по страновому российско-белорусскому признаку и являющуюся специфической для анализируемого массива. Проведенный анализ позволил также выявить макрогруппы объектов языка вражды во временном разрезе. Отметим, что этот результат, как и вообще обращение к динамическим аспектам процесса, были бы труднодоступными для исследования, опирающегося на реактивные методы. Полученные результаты указывают на возможность применения предлагаемой методики к широкому кругу текстовых материалов, а использование методов разведывательного анализа к обработке получаемых данных позволяет избежать ограничений, характерных для опросных инструментов.

Ключевые слова: поляризация, аффективная поляризация, язык вражды, язык ненависти, ВКонтакте, социальные сети

Для цитирования: *Стукал Д.К., Ахременко А.С., Петров А.П.* Аффективная политическая поляризация и язык ненависти: созданы друг для друга? // Вестник Российского университета дружбы народов. Серия: Политология. 2022. Т. 24. № 3. С. 480–498. https://doi.org/10.22363/2313-1438-2022-24-3-480-498

Благодарности: Исследование выполнено при финансовой поддержке РФФИ и АНО ЭИСИ в рамках научного проекта № 21-011-31340 «Политическая поляризация в России и Беларуси: эмпирический анализ „языка вражды“ в социальных медиа». Авторы благодарят за помощь в проведении исследования студентов НИУ ВШЭ: Д.А. Болатаеву, Д.А. Гончарову, Н.А. Деревцова, О.В. Золотову, К.А. Колесникову, А.В. Курбатова, Д.И. Левена, Е.А. Салтыкова, Н.С. Суханову, Д.Н. Чернова, А.Н. Шилину.

Affective Political Polarization and Hate Speech: Made for Each Other?

Denis K. Stukal[1], Andrei S. Akhremenko[1], Alexander P.C. Petrov[2]

[1] *National Research University Higher School of Economics, Moscow, Russian Federation*

[2] *Keldysh Institute for Applied Mathematics (Russian Academy of Sciences), Moscow, Russian Federation*

dstukal@hse.ru

Abstract. Abundant academic research has shown evidence of the growing polarization across the globe both in general and in terms of affective polarization. Previous research on this topic primarily employed reactive research methods like surveys or experiments, which however do not allow researchers to observe the behavior of the units of analysis in a natural setting. Presents an alternative approach that involves analyzing the observed behavior of social media users and identifying the key polarizing cleavages through the study of hate speech with respect to distinct target groups. We present a novel coding schema for textual data, which includes two components: first, an operationalized definition of hate speech as a phenomenon with at least one of the three elements — insult, discrimination, or aggression; and second, an original coding guide for human coders annotating the use of hate speech. We apply our approach to the analysis of empirical data that includes over 5000 posts on the social media platform VK about the meetings between the Presidents of Russia and Belarus in 2020–2021. After coding the collected data, we performed the empirical analysis that identified two generic cleavages. One is about domestic politics in Belarus and Russia, whereas the other is related to the opposition between these two countries on the one hand, and Western countries on the other. We also found an additional Russian/Belarusian cleavage that is peculiar to the collected dataset. Our methodology also allowed us to identify

and analyze the dynamics of macro-groups that were targets of hate speech. Importantly, these results — as any other dynamic aspect of analysis — would be highly challenging in research based on reactive methods. Thereby our results highlight the prospects of applying the proposed methodology to a broad range of textual data, as well as the benefits of exploratory analysis that helps overcome the limitations of survey instruments.

Keywords: polarization, affective polarization, hate speech, VK, social media

For citation: Stukal, D.K., Akhremenko, A.S., & Petrov, A.P. (2022). Affective political polarization and hate speech: Made for each other? *RUDN Journal of Political Science*, 24(3), 480–498. (In Russian). https://doi.org/10.22363/2313-1438-2022-24-3-480-498

Acknowledgements: The study was supported of the RFBR and ANO EISS in the framework of scientific project No. 21-011-31340 “Political Polarization in Russia and Belarus: an empirical analysis of the “hate speech” in social media”. The authors are grateful for the help of HSE students in conducting the study: D.A. Bolataeva, D.A. Goncharova, N.A. Derevtsov, O.V. Zolotova, K.A. Kolesnikov, A.V. Kurbatov, D.I. Leven, E.A. Saltykov, N.S. Sukhanova, D.N. Chernov, A.N. Shilina.

Аффективная поляризация и язык ненависти в современной политической науке

В политической науке последнего десятилетия ключевым трендом в исследовании феномена политической поляризации стал переход от одномерного представления, связанного с идеологией и ее проекциями на вопросы повестки дня, к гораздо более объемной и многомерной картине. Традиционным для политологии является понимание поляризации как расхождения позиций (positional polarization) между партиями и индивидами в некотором политическом измерении [Fiorina, Abrams 2008]: очень наглядным представлением здесь является центробежное движение по классической лево-правой шкале, восходящей еще к Великой французской революции. Разумеется, расхождение позиций может происходить и в иных широких ценностно-политических измерениях, и гораздо более локально — по какому-то вопросу текущей повестки или по отношению к определенной политической фигуре или группе. В последние годы позиционный взгляд дополняется еще как минимум двумя самостоятельными исследовательскими фокусами, связанными с концептами структурной и аффективной поляризации.

Понятие структурной поляризации (иногда называемой также поляризацией взаимодействий, interactional polarization [Yarchi, Baden, Kligler-Vilenchik 2021] сформировалось в рамках исследований цифровой политической коммуникации. Первоначальный взгляд на интернет как на среду, в наибольшей степени благоприятную конструктивному обмену различными точками зрения [Papacharissi 2002], по мере накопления исследовательских результатов менялся на почти противоположный. И сейчас для значительной части ученых один из ключевых отправных пунктов состоит в том, что особенности структурирования коммуникации, получения информации в рамках социальных медиа способствуют фрагментации поля дебатов, размежеванию политических позиций.

Это происходит за счет, во-первых, феномена «гомофилии» [McPherson, Smith-Lovin, Cook 2001] — склонности поддерживать общение с людьми, близкими по политическим взглядам, и дистанцироваться от идейных оппонентов [Settle 2018]. Именно в интернете, предоставляющем широкие возможности для установления / расторжения коммуникационных связей, по сравнению с «физическим» миром, и в целом для «фильтрации» контента (selective exposure [Bode 2016]), этот эффект наиболее силен. Во-вторых, доступность широкой палитры разнообразных точек зрения подталкивает многих пользователей социальных медиа скорее к поиску источников, подтверждающих их уже сложившиеся представления, нежели к участию в делиберациях с оппонентами [Wolleback et al. 2019]. Наконец, свой вклад вносят и алгоритмы рекомендательных и поисковых систем самих социальных медиа, учитывающих предшествующую активность пользователей и их социально-демографические характеристики [Cho et al. 2020]. В результате формируются почти закрытые от внешнего мира «эхо-камеры» (echo chambers) или «информационные пузыри» (filter bubbles), склонные концентрироваться на полюсах политических дискуссий [Bodrunova et al. 2019]. В этом, в очень кратком изложении, состоит суть и механизм структурной поляризации.

Третий тип политической поляризации, центральный для нашей работы, — аффективная поляризация (affective polarization). Она заключается в возрастании враждебности и агрессии по отношению к политическим оппонентам, а также разделяемым ими идеям и ценностям. При этом сами эти идеи и ценности могут не претерпевать никаких изменений. Примечательно, что авторы концепции аффективной политической поляризации[1] исходно рассматривали ее как сугубо северо-американский феномен, связанный с нарастающей неприязнью, вплоть до ненависти, между демократами и республиканцами [Iyengar, Sood, Lelkes 2012: 405]. Однако очень быстро концепт аффективной поляризации стал востребован исследователями самых различных страновых кейсов — от Израиля [Harel, Jameson, Maoz 2020] до Эфиопии [Gagliardone 2014] — и лег в основу самостоятельного и быстро развивающегося направления политических исследований.

Внимание к концепту аффективной поляризации обусловлено несколькими причинами, помимо явных эмпирических свидетельств в пользу ее значимости как фактора восприятия политики и коммуникаций в этой сфере [Iyengar et al. 2019].

Во-первых, это вписывается в наметившийся в последние годы тренд к большему учету эмоциональной составляющей принятия решений и эффектов внутри- и межгруппового влияния, в свою очередь ведущего к более широкому использованию в политической науке результатов социальной психологии. Именно эта дисциплина дает общую рамку рассмотрения аффективной поляризации, основанную на теории социальной идентичности (social identity theory) [Tajfel,

[1] «Программным», наиболее широко цитируемым текстом здесь является статья Ш. Ийенгара, Г. Соода и И. Лелкеса (из Стэндфордского, Принстонского и Амстердамского университетов соответственно) «Аффект, а не идеология», опубликованная в журнале Public Opinion Quarterly в 2012 г. [Iyengar, Sood, Lelkes 2012]

Turner 1979]. В соответствии с ними идентификация с некоторым — формальным или неформальным — сообществом (ингруппой) автоматически приводит к конструированию представления о противостоящей ему аутгруппе. Такое разделение порождает у индивида склонность положительно оценивать членов ингруппы и отрицательно — представителей аутгруппы только на основании воспринимаемой групповой принадлежности. Одно из важных следствий такого взгляда на механизм аффективной поляризации состоит в том, что это явление отражает размежевание *не только и не столько индивидов, сколько групп и сообществ*; мы будем делать на этом особый акцент в части описания эмпирического инструментария нашего исследования.

Во-вторых, аффективная поляризация в значительной мере связана с использованием цифровых каналов и платформ коммуникации, которые находятся в фокусе одного из наиболее динамично развивающихся направлений политической науки. Затронутые выше механизмы структурной поляризации — селективное восприятие, сетевая гомофилия, эхо-камеры — вполне могут быть одновременно и механизмами аффективной поляризации. В целом имеется большой потенциал для синтеза этих двух (как минимум) аспектов политической поляризации как на теоретическом, так и на эмпирическом уровне. Однако пока он реализован лишь в очень слабой мере; соответствующих работ на удивление мало (например: [Wolleback et al. 2019]).

Более того, «мейнстримом» эмпирических исследований аффективной поляризации остаются традиционные социологические методы. Частично это связано с американским происхождением самого концепта: именно в США оказались накоплены подходящие данные массовых социологических опросов. Так, в лонгитюдном проекте «Американское национальное исследование выборов» (American National Election Study, ANES) имеется вопрос-термометр относительно степени «теплоты» в отношении респондента к Республиканским и Демократическим партиям и их представителям: используется шкала от «холодного» (0) до «теплого» (100). Индивидуальная мера аффективной поляризации рассчитывается как расстояние между оценкой партии, к которой принадлежит респондент, и оценкой оппозиционной партии [Mason 2013].

Сходные социологические инструменты предполагают соотнесение респондентами представителей партий с этически и эмоционально окрашенными характеристиками: «доброжелательный», «умный», открытый», «щедрый» *vs* «эгоистичный», «злобный», «лицемерный» и т.п. [Iyengar, Sood, Lelkes 2012]. Еще один подход опирается на традиции исследований социального капитала и выводит меру аффективной поляризации из оценок степени доверия, испытываемого респондентами по отношению к оппозиционным политическим сообществам. В этом же русле лежит измерение социальной дистанции через оценку приемлемости и комфорта вступления в близкие личные отношения с оппонентами: например, отношение к перспективе брака ребенка с представителем оппозиционной партии [Iyengar et al. 2019; Druckman, Levendusky 2019].

Наряду с опросными методиками, основанными на самоотчетах (self-report), используются и поведенческие (behavioral) подходы, основанные на фиксации

наблюдаемых реакций. Основные экспериментальные техники, используемые для выявления и оценки аффективной поляризации, основываются на играх по распределению некоторого ограниченного объема финансовых средств — игра диктатора (dictator game), игра доверия (trust game) и т.п. [Carlin, Love 2013]. Уровень проявления аффективной поляризации измеряется здесь как разность в количестве денег, распределенных представителям своей группы по сравнению с «чужаками».

Характеризуя приведенные методики для выявления и оценки уровня аффективной поляризации, отметим, что все они являются реактивными: исследователь создает определенный стимул (в виде вопроса или правил игры) и оценивает реакцию респондента или испытуемого. Такая ситуация является искусственной в том смысле, что она может быть никак не связана с повседневным поведением и представлениями индивида. Будет ли естественное, невынужденное поведение соответствовать тем закономерностям, которые выявляют опросы и эксперименты? Однозначно положительного ответа на этот вопрос, вообще говоря, нет. В случае с разными формами самоотчета проблема усугубляется эффектом социальной желательности.

Отметим и более технический, но важный момент: с помощью экспериментальных техник и опросов со сложным дизайном трудно набирать длинные временные ряды наблюдений. По данной тематике доступны лишь лонгитюдные исследования с простейшими вопросами вроде «термометров» в названном ANEC, которые исходно не планировались как инструменты измерения собственно аффективной поляризации. А наличие рядов динамики очень важно хотя бы потому, что само понятие поляризации отражает не только явление, но и процесс.

Наконец, опросные методы хороши, когда известно, какие вопросы следует задавать. Другими словами, когда понятны ключевые линии политических размежеваний и те группы, которые противостоят друг другу. В отличие от США, где политический процесс четко структурирован вокруг борьбы двух ведущих партий, в России (и многих других странах) это совсем не очевидно.

Мы полагаем, что указанные проблемы способен преодолеть подход, основанный на фиксации элементов языка вражды (hate speech) в онлайн-дискурсе вокруг значимых политических событий. Язык вражды (ненависти[2]) мы пока в общем виде определим как вербальное поведение, направленное на унижение достоинства, в том числе через призывы к насилию или дискриминации[3]. Таким образом, язык вражды может рассматриваться как наблюдаемое проявление аффективной поляризации уже как минимум в том отношении, что он отражает ее ключевую эмоциональную (собственно аффективную) составляющую — враждебность и агрессию по отношению к оппонентам.

[2] В данной работе понятия «язык вражды» и «язык ненависти» употребляются как синонимы.

[3] Детальное операциональное определение будет дано в разделе, посвященном авторской методике.

Однако эти явления родственны и на более глубоком теоретическом уровне. В современной литературе одним из ключевых признаков языка вражды является адресация к групповой принадлежности объекта ненависти [Olteanu et al. 2018; Siegel 2020; Kennedy et al. 2018]. Так, оскорбительные высказывания или даже призывы к насилию в отношении конкретного индивида не определяются как язык вражды, если нет явной или скрытой отсылки к целевой группе (target group); достоинство индивида подвергается унижению в силу его или ее принадлежности к некоторой социальной общности, которая и является действительным объектом ненависти. Та же логика, как мы отмечали выше, справедлива и применительно к аффективной поляризации, понимаемой в русле теории социальной идентичности: размежевание проходит на уровне групп и сообществ, а не персоналий.

С эмпирической точки зрения язык ненависти также выглядит естественным индикатором аффективной поляризации. Это явно наблюдаемое поведение, регистрируемое по набору вполне операциональных признаков. При этом исследователь не создает контролируемой ситуации с заданными вариантами поведения: выборка сообщений формируется из возникающей естественным путем совокупности. Также не ожидается каких-либо форм самоотчета со стороны обследуемых. Использование данных социальных медиа позволяет сравнительно легко решить проблемы сбора ретроспективной информации и формирования рядов динамики. Наконец, огромные массивы данных интернет-коммуникаций позволяют оптимально организовать поисковое исследование: определить неочевидные заранее линии аффективной поляризации и идентифицировать противоборствующие группы.

Однако в политической науке мы располагаем лишь сравнительно небольшим числом исследований, в которых аффективная поляризация и язык вражды используются в прямой связке (см., например: [Harel, Jameson, Maoz 2020; Kennedy et al. 2018]). На фоне сказанного выше это выглядит почти парадоксальным.

Мы полагаем, что дело в своего рода инерции исследовательских традиций, сложившихся независимо друг от друга в изучении аффективной политической поляризации и языка вражды. Последний еще с 1990-х гг. рассматривается почти исключительно в контексте отношения к расовым, этническим, религиозным, сексуальным меньшинствам. Причем такой взгляд сформировался не только у исследователей (например: [Jacobs, Potter 1998]), но и на уровне официальных организаций — вплоть до Европейского союза[4]. Такой фокус фактически оставляет «за бортом» проблематики аффективной поляризации расколы по собственно политическим линиям. В исследованиях же аффективной поляризации наметилась своя «колея», связанная с изучением партийной поляризации, прежде всего в США.

[4] Council of Europe (1997). Recommendation No. R (97) 20 of the Committee of Ministers to Member States on “hate speech”. Доступно по ссылке: URL: https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680505d5b

В результате подход к исследованию аффективной политической поляризации на основе анализа языка ненависти в социальных медиа представляется нам существенно недооцененным. Далее мы покажем, как созданная авторами методика работает на материале сообщений в социальной сети ВКонтакте, организованных вокруг встреч президентов В. Путина и А. Лукашенко в 2020–2021 гг.

Выявление языка ненависти в социальных медиа

Методики выявления и кодирования языка ненависти в социальных медиа различаются в нескольких измерениях. Так, различаются те из них, что проводятся с использованием ручного кодирования, и те, что используют лишь машинную обработку (при этом обучение алгоритма может проводиться с использованием ручных процедур). В настоящей работе формирование выборки сообщений, содержащих язык ненависти, происходит в два этапа: машинный сбор данных по тематическому принципу и ручное выявление таких сообщений, которое происходит в рамках той же процедуры, что и кодирование, главной целью которого является классификация таких сообщений. Сходным образом поступили, в частности, авторы работы [Gitari et al. 2015]. На первом этапе они провели первоначальный отбор сообщений с так называемых «сайтов ненависти», а на втором классифицировали их как содержащие сильно выраженный, слабо выраженный и не содержащие язык вражды (strongly hateful, weakly hateful, non-hateful).

Другой подход (в частности [Olteanu et al. 2018]) предполагает, что данные уже изначально собираются так, чтобы выборка содержала лишь сообщения с языком ненависти. Это достигается за счет того, что сбор сообщений идет по ключевым словам, присущим языку вражды. Фундаментальная проблема состоит в том, что как ручной анализ, так и, тем более, машинный не всегда позволяют удовлетворительно решить вопрос о наличии языка вражды в том или ином сообщении. В целом более «машинные» процедуры ориентированы на высокую чувствительность (high recall, high sensitivity, собрать как можно больше сообщений, содержащих язык вражды), а более «ручные» — на высокую специфичность (high precision, high specificity, избежать попадания в выборку сообщений, не содержащих его).

Другой аспект процедур отбора и кодирования связан с тем, что, в зависимости от цели исследования, некоторые методики ориентированы на выявление языка вражды в отношении лишь некоторых групп как объектов ненависти, другие методики носят в этом плане универсальный характер. Например, при изучении языка вражды в адрес мусульман в качестве ключевых слов используются специфические ключевые слова, характерные для исламофобских групп [Olteanu et al. 2018]. В противоположность этому, в работе [Kennedy et al. 2018] формирование выборки происходит по алгоритму, допускающему различные группы в качестве объекта ненависти, а частью анализа было кодирование, определяющее конкретную группу: направлена ли нена-

висть данного сообщения на группу по признаку расы (этничности), национальности (в смысле принадлежности к той или иной стране, nationality), пола, религии, сексуальной ориентации, идеологии, политической (партийной) ориентации, состояния ментального или физического здоровья.

Наконец, отметим еще один вопрос, специфический для эмпирических исследований данной области. Высказывание считается содержащим язык вражды, лишь если объект ненависти определяется по признаку принадлежности к определенной социальной группе. Другими словами, если ненависть направлена на индивида как такового, то это не считается языком вражды. При кодировании возникает вопрос о том, каким образом определять наличие групповой направленности. Проблема не возникает, если указание на группу содержится в самом агрессивном высказывании. Однако иногда такое указание отсутствует в агрессивном высказывании, но содержится в более широком контексте. Относить ли такие случаи к языку вражды? В случае положительного ответа, каким образом возможно формализовать контекст, чтобы критерий имел четкую форму?

Покажем, как может быть формализован контекст, на примере работы [Olteanu et al. 2018], в которой под языком вражды понимаются высказывания, которые могут быть восприняты как оскорбительные, унизительные или каким-либо образом вредные и которые мотивированы, полностью или частично, чьей-либо предвзятостью в отношении какого-либо аспекта группы людей, либо комментариями такой речи другими людьми, либо речью, направленной на противодействие языку вражды. Проиллюстрируем понимание контекста, принятое в данном определении, с помощью следующего примера.

Пользователь A: Оскорбление в адрес какой-либо социальной группы. (1)

Пользователь B: Ответное оскорбление в адрес A, не называющее его групповую принадлежность. (2)

Пользователь A: Ответное оскорбление в адрес B, не называющее его групповую принадлежность. (3)

Пользователь C: Оскорбление в адрес B, не называющее его групповую принадлежность. (4)

В данном примере очевидно, что высказывание (1) содержит язык вражды, а высказывание (2) — не содержит. Высказывания (3) и (4) не содержат указания на групповую принадлежность объекта ненависти (т.е. пользователя B), однако мотивированы его высказыванием (2), направленным на противодействие высказыванию (1), содержащему язык вражды. Следовательно, (3) и (4) также являются проявлениями языка вражды.

В противоположность использованному здесь определению, если исследовательский подход требует, чтобы указание на групповую принадлежность объекта ненависти содержалось в самом агрессивном высказывании (а не в контексте), то (3) и (4) не признаются языком вражды. Заметим, что обращение к контексту существенно повышает техническую сложность работы, так как требует анализа не только отдельных сообщений, но также связей между ними. Очевидно, методики, принимающие групповую направленность ненависти, содержащуюся

в контексте, относят к языку вражды большее количество сообщений, чем методики, требующие указания на группу в самом высказывании. Таким образом, здесь также имеет место указанная выше дилемма между чувствительностью и специфичностью.

Методология исследования

В данной работе осуществлен машинный сбор данных на основе тематического принципа, а кодировка, выделяющая из общего массива сообщения, содержащие язык ненависти, проведена вручную.

Исследование проводилось на материале публикаций (постов и комментариев) пользователей социальной сети ВКонтакте в связи с тематикой российско-белорусских отношений, более конкретно — в связи с переговорами президентов В.В. Путина и А.Г. Лукашенко, прошедшими в сентябре 2020 г., а также феврале, мае и сентябре 2021 г.

В соответствии с данной тематикой с помощью API ВКонтакте был собран корпус постов, в которых одновременно упоминались оба Президента, а также всех доступных комментариев к ним. Как правило, посты представляли собой информационные сообщения, например: «Лукашенко передал Бастрыкину привет от Путина и рассказал, о чем они до ночи говорили в Кремле» (и ссылка на публикацию в издании масс-медиа). В то же время комментарии не были сконцентрированы на обсуждении переговоров. По большей части они были посвящены обсуждению российско-белорусских отношений в общем смысле; но соизмеримое место занимал обмен мнениями и оценками по самому широкому кругу политических, идеологических и смежных с ними вопросов (включая, например, расстрел царской семьи, политическую ситуацию в Афганистане и т.д.).

Для ручной кодировки и анализа из этого корпуса методом случайной выборки был выбран 101 пост. Суммарно эти посты содержали 5503 комментария пользователей. Распределение собранных данных по времени представлено в табл. 1.

Таблица 1

Распределение комментариев из выборки по времени, 2020–2021

Месяц, год	Число постов	Общее число комментариев	Минимальное число комментариев к посту	Среднее число комментариев к посту	Максимальное число комментариев к посту
Сентябрь 2020	20	1100	1	55.0	100
Февраль 2021	14	801	18	57.2	100
Май 2021	17	951	20	55.9	100
Сентябрь 2021	50	2651	20	53.0	100

Примечание. Максимальное число комментариев к посту (100) обусловлено ограничениями API VK.
Источник: составлено авторами по результатам исследования.

Table 1

Distribution of comments from the sample by time, 2020–2021

Month, Year	Number of posts	Total number of comments	Min number of comments to the post	Mean number of comments to the post	Max number of comments to the post
September 2020	20	1100	1	55.0	100
February 2021	14	801	18	57.2	100
May 2021	17	951	20	55.9	100
September 2021	50	2651	20	53.0	100

Note. the maximum number of comments to a post (100) is due to restrictions of API VK.
Source: compiled by the authors based on the results of the study.

Отобранные таким образом комментарии были закодированы вручную в соответствии с разработанным в рамках данного исследования протоколом. Цель кодирования состояла в выявлении признаков языка ненависти и классификации комментариев в соответствии с этими признаками, а также оценки дополнительных переменных, необходимых для анализа (см. ниже).

Определение языка вражды, принятое в данной работе, имеет следующий вид. *Язык вражды — это агрессивные, оскорбительные и/или дискриминационные высказывания в отношении человека или группы лиц из-за их социальной, политической, национальной, этнической, религиозной, гендерной или иной принадлежности / идентичности*.

Соответственно, относительно конкретного комментария протокол кодировки требует указать наличие или отсутствие следующих признаков:

- призыв к насилию или угроза применения насилия в отношении человека или группы лиц из-за их принадлежности к той или иной группе;
- оскорбление в отношении человека или группы лиц из-за их принадлежности к той или иной группе;
- призыв к дискриминации или угроза дискриминации человека или группы лиц из-за их принадлежности к той или иной группе;
- является ли комментарий спамом / оффтопиком;
- выражает ли комментарий прямое несогласие — например, с исходным постом или другими комментариями;
- являются ли насилие, оскорбление или дискриминация основным посылом комментария.

Кроме того, в случае наличия признаков насилия, оскорбления или дискриминации кодировщикам следовало указать, кто/что является объектом ненависти (открытый вопрос, краткая формулировка).

Каждый комментарий рассматривался двумя либо тремя кодировщиками. Наличие признака фиксировалось, если его указали не менее половины кодировщиков (т.е. не менее одного из двух либо не менее двух из трех).

Всего в исследовании участвовало 11 кодировщиков, в качестве которых выступали студенты НИУ ВШЭ. Нагрузка на каждого кодировщика составляла примерно 510 комментариев в неделю в домашних условиях. Перед выполнени-

ем задания проводился групповой онлайновый инструктаж, в ходе которого кодировщики были ознакомлены с заданием. Был проведен тренировочный раунд кодирования с последующим разбором наиболее сложных кейсов. Во время первичного инструктажа, а также обсуждения результатов тренировочного раунда особое внимание кодировщиков было обращено на то, что языком вражды считаются лишь те проявления агрессии, оскорбления и дискриминация, которые направлены на индивидов ввиду их групповой принадлежности (применительно к данной работе чаще всего — ввиду принадлежности к числу сторонников того или иного политического направления). При том, что существуют различные подходы к тому, каким образом определять наличие указания на групповую принадлежность, в данной работе принят узкий критерий, требующий, чтобы указание на групповую принадлежность объекта языка вражды содержалось в том же комментарии, что и агрессия / оскорбление / дискриминация.

Результаты обработки комментариев всеми кодировщиками сводились в таблицу Excel с последующим статистическим анализом полученных данных методами описательной статистики.

Эмпирические результаты

Описательные статистики, характеризующие результаты кодирования собранных данных, представлены в табл. 2. Первая строка носит справочный характер и отражает общее число комментариев в данный период. Строки 2–4 содержат число комментариев с одним из трех типов языка вражды, а также процент таких комментариев от общего числа комментариев в данный период. Последняя строка является обобщением трех предыдущих строк и указывает на число комментариев с любым из трех признаков языка вражды. Поскольку один и тот же комментарий может иметь сразу несколько признаков языка вражды, число таких комментариев необязательно является суммой строк 2–4.

Как видно из данных табл. 2, язык вражды по анализируемой тематике выражается преимущественно, в оскорбительных комментариях. Комментарии дискриминирующего характера существенно уступают оскорблениям по частоте, в то время как угрозы насилия встречаются крайне редко. В целом частотность комментариев с признаками языка вражды оставалась относительно стабильной на протяжении рассматриваемого года, с небольшим ростом в середине 2021 г. (с 7 до 10–11 %).

Наибольший интерес в данном случае представляют не числовые характеристики динамики, а содержательные аспекты использования языка вражды. В ходе анализа полученных в результате кодирования данных была проведена классификация групп — объектов языка вражды. Были выделены шесть макрогрупп:

1) национально-территориальная;
2) политико-идеологическая;
3) профессиональная;
4) возрастная;
5) гендерная;
6) ЛГБТ;

Таблица 2

Описательные статистики результатов кодирования комментариев, 2020–2021

Число комментариев	Сент. 2020	Февр. 2021	Май 2021	Сент. 2021	ИТОГО
Общее число комментариев	1100	801	951	2651	5503
Число комментариев с оскорблениями	66 (6 %)	48 (6 %)	96 (10 %)	252 (10 %)	462 (8 %)
Число комментариев с угрозой насилия	5 (<0.5 %)	1 (<0.5 %)	11 (1 %)	17 (1 %)	34 (1 %)
Число комментариев с дискриминацией	20 (2 %)	17 (2 %)	31 (3 %)	23 (1 %)	91 (2 %)
Общее число комментариев с признаками языка вражды	72 (7 %)	55 (7 %)	107 (11 %)	275 (10 %)	509 (9 %)

Примечание. К числу комментариев с оскорблениями, угрозами насилия, дискриминацией — в соответствии с определением языка вражды — относились лишь комментарии, содержащие высказывания в отношении человека или группы лиц из-за их социальной, политической, национальной, этнической, религиозной, гендерной или иной принадлежности / идентичности. Последняя строка (число комментариев с признаками языка вражды) может не являться суммой комментариев с оскорблениями, угрозой насилия или дискриминацией, поскольку эти разновидности языка вражды могут одновременно встречаться в одном комментарии.

Источник: составлено авторами по результатам исследования

Table 2

Descriptive statistics of the results of coding comments, 2020–2021

Number of comments	September 2020	February 2021	May 2021	September 2021	Total
Total number of comments	1100	801	951	2651	5503
The number of comments with insults	66 (6 %)	48 (6 %)	96 (10 %)	252 (10 %)	462 (8 %)
The number of comments with the threat of violence	5 (<0.5 %)	1 (<0.5 %)	11 (1 %)	17 (1 %)	34 (1 %)
Number of comments with discrimination	20 (2 %)	17 (2 %)	31 (3 %)	23 (1 %)	91 (2 %)
Total number of comments with signs of hate speech	72 (7 %)	55 (7 %)	107 (11 %)	275 (10 %)	509 (9 %)

Note. Comments with insults, threats of violence, discrimination — in following the definition of the hate speech — included only comments containing statements against a person or group of persons because of their social, political, national, ethnic, religious, gender or other affiliation/identity. The last line (the number of comments with signs of hate speech) may not be the sum of comments with insults, threats of violence or discrimination, since these varieties of hate speech may occur simultaneously with one comment.

Source: compiled by the authors based on the results of the study.

К национально-территориальной макрогруппе относятся различные национальные (американцы, белорусы, русские, украинцы и др.), расовые (афроамериканцы), региональные (выходцы с Донбасса, европейцы) группы. К политико-идеологической макрогруппе относятся политические элиты двух стран, группы сторонников и противников действующей власти в России и Беларуси, идеологические группы (либералы, консерваторы, коммунисты). Профессиональная макрогруппа образована в основном тремя группами: силовики, чиновники и мигранты. Возрастная макрогруппа состоит из молодежи и пожилых; гендерная и ЛГБТ — соответственно названию.

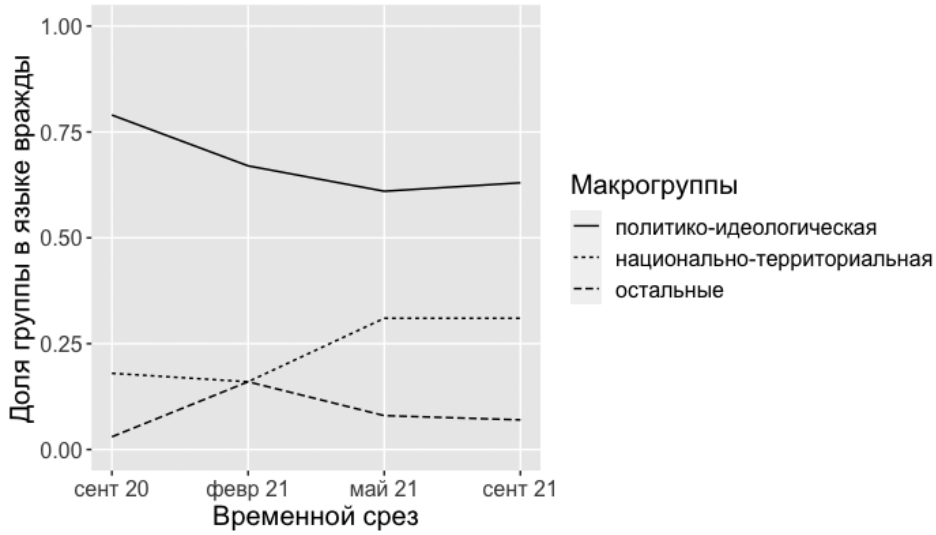

Рис. 1. Динамика долей макрогрупп объектов языка вражды
Источник: составлено авторами по результатам исследования.

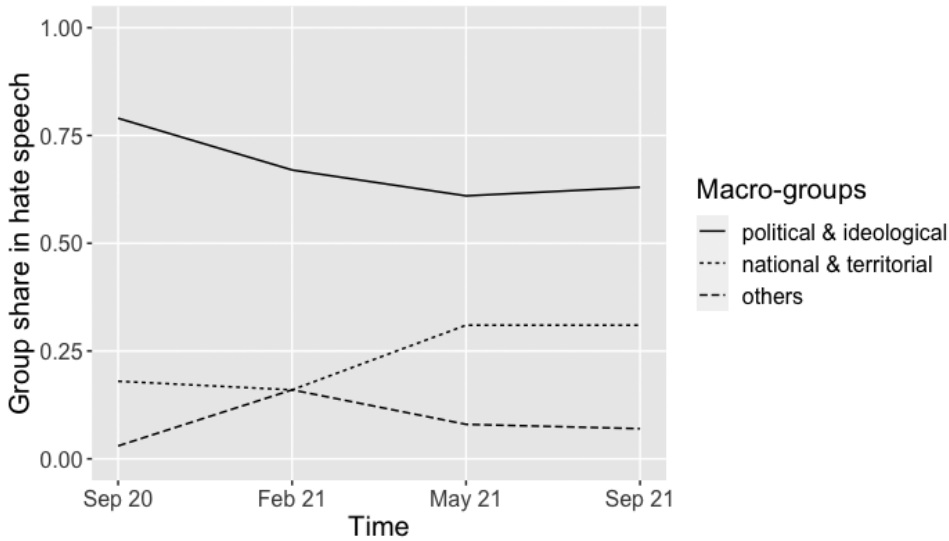

Fig. 1. Dynamics of the shares of macro groups of objects of the hate speech
Source: compiled by the authors based on the results of the study.

Данные рис. 1 показывают, как доли этих макрогрупп объектов языка вражды менялись во времени. Как видно из графика, доминирующее положение занимает политико-идеологическая макрогруппа, на долю которой приходилось от 60 до 80 % случаев использования языка вражды. Преобладание этой макрогруппы объектов неудивительно, поскольку тематика отобранных постов носила политический характер, в связи с чем политические позиции героев сообщений или участников обсуждений становились объектом оскорблений или других форм языка вражды.

Более интересна вторая по частотности, национально-территориальная, макрогруппа, на долю которой приходится около 25 % случаев использования языка вражды. Достаточно частое появление этой макрогруппы в контексте тематики анализируемых постов неудивительно: встречи двух Президентов были связаны с темой углубления интеграции двух государств и развития Союзного государства, что могло спровоцировать апелляцию к национальным и территориальным сюжетам (табл. 3). Интерес представляет другое. Во-первых, обращает на себя внимание динамика доли этой макрогруппы: она демонстрирует явный скачок в середине 2021 г., когда стали понятны контуры планируемых мер по дальнейшей интеграции. Во-вторых, любопытен тот факт, что наиболее часто фигурирующей в языке вражды национальной группой являются не русские и белорусы, а украинцы, упоминаемость которых в 2–3 раза превышает упоминаемость русских и белорусов. Не единичны также случаи упоминания американцев или европейцев — часто в контексте оскорбительных или дискриминационных высказываний в отношении ЛГБТ. Эти закономерности указывают на достаточно прочную укорененность ряда шаблонов (связка США/Европа — ЛГБТ) и повесток (упоминания США при обсуждении встреч В.В. Путина и А.Г. Лукашенко) в сознании участников онлайн-обсуждений.

Таблица 3

Группы объектов языка вражды с высокой частотностью

Политико-идеологическая макрогруппа		**Национально-территориальная макрогруппа**	
Руководство стран	178 (58 %)	Украинцы	55 (43 %)
Оппозиция	57 (19 %)	Русские / россияне	25 (19 %)
Сторонники власти	54 (18 %)	Белорусы	20 (16 %)
Либералы	15 (5 %)	Евреи	8 (6 %)
		Американцы	4 (3 %)
		Европейцы	3 (2 %)

Примечание. Проценты рассчитаны от числа комментариев, относящихся к данной макрогруппе объектов. Таблица содержит лишь одну идеологическую группу («либералы»), поскольку другие идеологические группы встречаются редко: «коммунисты» — 2 раза, «консерваторы» — 1 раз.

Источник: составлено авторами по результатам исследования.

Table 3

Groups of objects of hate speech with high frequency

Political and ideological macrogroup		National-territorial macrogroup	
Country leadership	178 (58 %)	Ukranians	55 (43 %)
Opposition	57 (19 %)	Russians	25 (19 %)
Supporters of the government	54 (18 %)	Belarusians	20 (16 %)
Liberals	15 (5 %)	Jews	8 (6 %)
Other ideological groups are rare		Americans	4 (3 %)
		Europeans	3 (2 %)

Note. Percentages are calculated from the number of comments related to this macro group of objects. The table contains only one ideological group ("liberals"), since other ideological groups are rare: "communists" — twice, "conservatives" — once.

Source: compiled by the authors based on the results of the study.

Высокая встречаемость политико-идеологической и национально-территориальной макрогрупп объектов языка вражды, а также частотность конкретных групп внутри выделенных макрогрупп (см. табл. 3) позволяет выявить несколько ключевых расколов, возникших в онлайн-сообществах ВКонтакте в контексте обсуждения встреч В.В. Путина и А.Г. Лукашенко.

Первый раскол пролегает по линии отношения к действующей власти: лоялисты — оппозиция. Можно утверждать, что именно этот раскол обладает наибольшей эмоциональной окраской и характеризуется наибольшей аффективной поляризацией в политической онлайн-коммуникации. При этом объектами языка вражды выступают как сторонники и противники власти, так и сами представители власти. Показательно, что последние являются наиболее частотной группой среди объектов языка вражды. Общий характер данного раскола, не связанный с тематикой отобранных сообщений, позволяет предполагать, что выявленный раскол носит устойчивый системный характер и выходит за рамки тематической выборки, анализируемой в данном проекте.

Второй раскол, также претендующий на статус системного, пролегает по линии противопоставления России/Беларуси странам Запада. В рамках этого национально-территориального раскола зачастую актуализируется также тематика ЛГБТ-сообществ, придающая расколу не только национальный или политический, но и ценностный характер.

Наконец, специфическим для анализируемого массива эмпирических данных расколом является противопоставление россиян и белорусов. Конкретные формы проявления этого раскола достаточно разнообразны и включают в себя экономические факторы (от негативного отношения к российским предпринимателям до обвинения белорусов в намерении получить дотации из российского бюджета), геополитических аллюзий («быть польскими холопами») или их смесью («Прощай Белаз — здравствуй польский унитаз!»). Достаточно часто данный раскол возникал также параллельно расколу по линии отношения к действующей власти и выражал негативное отношение комментаторов, поддерживающих власть/оппозицию в одной стране, к комментаторам, поддерживающим противоположный лагерь в другой стране.

Таким образом, анализ эмпирических данных, характеризующих политическую онлайн-коммуникацию по достаточно узкой тематике встреч двух президентов, позволил выявить две общие линии политического раскола и дополнить их более специфическим, тематически обусловленным расколом.

Заключение

Несмотря на то, что исследования аффективной поляризации переживают в последние годы бум, методология этих исследований до сих пор находится на ранних стадиях своего развития. Исследователи в значительной мере остаются ограниченными теми инструментами реактивного характера, которые доступны им благодаря межстрановым социологическим опросам, в первую очередь Сравнительное исследование избирательных систем (Comparative Study of Electoral Systems, CSES). Несмотря на преимущества использования сопоставимых в страновом и временном разрезах данных, опора на подобные стандартизированные инструменты имеет и ряд недостатков, связанных с низкой адаптивностью к быстро меняющемуся политическому контексту либо уникальным особенностями конкретного странового контекста, а также с зависимостью исследователей от конкретных вопросов, сформулированных группой разработчиков опроса.

В данной статье мы предлагаем альтернативную — нереактивную — методологию исследования, опирающуюся не на опросы общественного мнения или экспериментальные методы, а на разведывательный анализ поведенческих данных текстового характера. С опорой на существующую литературу мы разработали авторскую методику кодирования языка вражды, позволяющую выявлять как факты использования языка вражды в отношении тех или иных групп, так и ее разновидности. Далее мы предлагаем выявлять актуальные линии раскола, характеризующиеся высоким уровнем аффективной поляризации, на основании выявляемых в онлайн-коммуникации наиболее частотных групп объектов языка вражды. Данная методика апробируется нами на эмпирическом материале большого числа русскоязычных комментариев пользователей социальной сети ВКонтакте, посвященных встречам Президентов России и Беларуси в сентябре 2020 — сентябре 2021 г.

Проведенная нами эмпирическая апробация предлагаемой методики указывает на высокий аналитический потенциал последней: в массиве нескольких тысяч комментариев были выявлены две устойчивые линии раскола, связанные с отношением к действующей власти, а также противопоставлением России/Беларуси странам Запада, а также достаточно специфическая для анализируемого массива линия раскола по страновому российско-белорусскому признаку.

Предлагаемая методика может быть применена к исследованию более широкого круга вопросов и политико-страновых контекстов. Действительно, предлагаемая нами методика кодирования может применяться к разнообразным текстам на различных языках, охватывающим произвольно широкий круг вопросов, а разведывательный характер анализа данных о выявленных в текстах группах объектов языка вражды позволяет избежать жестких априорных шаблонов вопросов, неизбежных при использовании опросных методик. При этом предла-

гаемая нами методика может выступать как субститутом, так и комплементом опросных методик: в одном варианте наша методика может использоваться на ранних стадиях исследования и служить основой для разработки контекстуально-информированных опросов; в другом варианте эмпирическое исследование, опирающееся на предлагаемую нами методику, может использоваться для насыщения фактурой результатов проведенных опросов.

Совмещение обеих методик эмпирического исследования, опирающихся на реактивные и нереактивные методы, открывает новые возможности для исследования как аффективной поляризации, так и в целом политического поведения в сравнительной перспективе.

Поступила в редакцию / Received: 30.01.2022
Доработана после рецензирования / Revised: 01.06.2022
Принята к публикации / Accepted: 15.06.2022

References / Библиографический список

Bode, L. (2016). Pruning the news feed: Unfriending and unfollowing political content on social media. *Research & Politics*, July 2016, 1–8. https://doi.org/10.1177/2053168016661873

Bodrunova, 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. https://doi.org/10.17645/mac.v7i3.1934

Carlin, R.E., & Love, G.J. (2013). The politics of interpersonal trust and reciprocity: An experimental approach. *Political Behavior*, 35(1), 43–63. https://doi.org/10.1007/s11109-011-9181-x

Cho, J., Ahmed, S., Hilbert, M., Liu, B., & Luu, J. (2020). Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization. *Journal of Broadcasting & Electronic Media*, 64(2), 150–172. https://doi.org/10.1080/08838151.2020.1757365

Druckman, J., & Levendusky, M. (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly*, 83(1), 114–122. https://doi.org/10.1093/poq/nfz003

Fiorina, M.P., & Abrams, S.J. (2008). Political polarization in the American public. *Annual Review of Political Science*, 11(1), 563–588. https://doi.org/10.1146/annurev.polisci.11.053106.153836

Gagliardone, I. (2014). Mapping and analysing hate speech online. Retrieved April 24, 2022 from *SSRN*: https://ssrn.com/abstract=2601792

Gitari, N.D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering*, 10(4), 215–230. https://doi.org/10.14257/ijmue.2015.10.4.21

Harel, T.O., Jameson, J.K., & Maoz, I. (2020). The normalization of hatred: Identity, affective polarization, and dehumanization on Facebook[5] in the context of intractable political conflict. *Social Media + Society*, April–June, 1–10. https://doi.org/10.1177/2056305120913983

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S.J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: Social identity perspective on polarization. *Public opinion quarterly*, 76(3), 405–431. https://doi.org/10.1093/poq/nfs038

[5] 21 марта 2022 г. Тверской суд города Москвы признал Meta (продукты Facebook и Instagram) экстремистской организацией.

Jacobs, J., & Potter, K. (1998). *Hate crimes: Criminal law and identity politics*. New York, NY: Oxford University Press.

Kennedy, B., Atari, M., Davani, A.M., Yeh, L., Omrani, A., Kim, Y., Coombs, K., Havaldar, S., Portillo-Wightman, G., Gonzalez, E., & Hoover, J. (2018). The Gab Hate Corpus: A collection of 27k posts annotated for hate speech. *PsyArXiv Preprint*. Retrieved April 24, 2022, from https://psyarxiv.com/hqjxn/

Mason, L. (2013). The rise of uncivil agreement: Issue versus behavioral polarization in the American electorate. *American Behavioral Scientist*, 57(1), 140–159. https://doi.org/10.1177/0002764212463363

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27, 415—444. https://doi.org/10.1146/annurev.soc.27.1.415

Olteanu, A., Castillo, C., Boy J., & Varshney K. (2018). The effect of extremist violence on hateful speech online. *arXiv preprint*. Retrieved April 24, 2022, from arXiv:1804.05704

Papacharissi, Z. (2002). The virtual sphere: The internet as a public sphere. *New Media & Society*, 4(1), 9–27. https://doi.org/10.1177/14614440222226244

Settle, J.E. (2018). *Frenemies: how social media polarizes America*. Cambridge, New York: Cambridge University Press.

Siegel, A. (2020). Online Hate Speech. In N. Persily & J. Tucker (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform* (56-88). Cambridge: Cambridge University Press.

Tajfel, H., & Turner, J.C. (1979). An integrative theory of intergroup conflict. In W.G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (33-37). Monterey, CA: Brooks/Cole.

Wolleback, D., Karlsen, R., Steen-Johnsen, K., & Enjolras, B. (2019). Anger, fear, and echo chambers: The emotional basis for online behavior. *Social Media + Society*, 5(2), 1–14. https://doi.org/10.1177/2056305119829859

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, 38(1-2), 98–139. https://doi.org/10.1080/10584609.2020.1785067

Сведения об авторах:

Стукал Денис Константинович — кандидат политических наук, PhD, ведущий научный сотрудник Института прикладных политических исследований, Национальный исследовательский университет „Высшая школа экономики“ (e-mail: dstukal@hse.ru) (ORCID: 0000-0001-6240-5714)

Ахременко Андрей Сергеевич — доктор политических наук, профессор факультета социальных наук, Национальный исследовательский университет «Высшая школа экономики» (e-mail: aakhremenko@hse.ru) (ORCID: 0000-0001-8002-7307)

Петров Александр Пхоун Чжо — доктор физико-математических, ведущий научный сотрудник Института прикладной математики имени М.В. Келдыша РАН (e-mail: petrov.alexander.p@yandex.ru) (ORCID: 0000-0001-5244-8286)

About the authors:

Denis K. Stukal — Cand. Sci. (Pol. Sci.), PhD, Leading Research Fellow, Institute for Applied Political Studies, HSE University, Moscow (e-mail: dstukal@hse.ru) (ORCID: 0000-0001-6240-5714)

Andrei S. Akhremenko — Dr. Sci. (Pol. Sci.), Professor, School of Social Sciences, HSE University (e-mail: aakhremenko@hse.ru) (ORCID: 0000-0001-8002-7307)

Alexander P.C. Petrov — Dr. Sci. (Applied Math.), Senior Researcher, Keldysh Institute for Applied Mathematics (Russian Academy of Sciences) (e-mail: petrov.alexander.p@yandex.ru) (ORCID: 0000-0001-5244-8286)