NBER WORKING PAPER SERIES

SOCIAL MEDIA AND XENOPHOBIA: EVIDENCE FROM RUSSIA

Leonardo Bursztyn
Georgy Egorov
Ruben Enikolopov
Maria Petrova

Working Paper 26567
http://www.nber.org/papers/w26567

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2019

We are grateful to Matthew Gentzkow, Brian Knight, Aakaash Rao, Alvaro Sandroni, David Strömberg, and Alireza Tahbaz-Salehi for very helpful discussions, to Danil Fedchenko for excellent research assistance, and to numerous seminar participants for comments and suggestions. Ruben Enikolopov and Maria Petrova acknowledge financial support from the BBVA foundation grant. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Social Media and Xenophobia: Evidence from Russia
Leonardo Bursztyn, Georgy Egorov, Ruben Enikolopov, and Maria Petrova
NBER Working Paper No. 26567
December 2019
JEL No. D7,H0,J15

ABSTRACT

We study the causal effect of social media on ethnic hate crimes and xenophobic attitudes in Russia using quasi-exogenous variation in social media penetration across cities. Higher penetration of social media led to more ethnic hate crimes, but only in cities with a high pre-existing level of nationalist sentiment. Consistent with a mechanism of coordination of crimes, the effects are stronger for crimes with multiple perpetrators. We implement a national survey experiment and show that social media persuaded young and low-educated individuals to hold more xenophobic attitudes, but did not increase respondents' openness to expressing these views. Our results are consistent with a simple model of social learning where penetration of social networks increases individuals' propensity to meet like-minded people.

Leonardo Bursztyn
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and NBER
bursztyn@uchicago.edu

Georgy Egorov
Kellogg School of Management
Northwestern University
2211 Campus Drive
Evanston, IL 60208
and NBER
g-egorov@kellogg.northwestern.edu

Ruben Enikolopov
Barcelona IPEG
Universitat Pompeu Fabra
Barcelona GSE, New Economic School
Edif. Mercè Rodoreda 23.105 (IPEG)
C/ Ramon Trias Fargas, 25-27
08005 Barcelona, Spain
ruben.enikolopov@upf.edu

Maria Petrova
Barcelona IPEG
Ramon Trias Fargas 25-27
Barcelona
Spain
petrova.ma@gmail.com

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/3066

1 Introduction

In recent years, the world has witnessed a large increase in expression of hate and xenophobia.[1] Candidates and platforms endorsing nationalism and views associated with intolerance toward specific groups have also gathered increased popular support both in the U.S. and across Europe. Social media has been widely named a major factor in the increase in expression of hate, and hate crimes in particular.[2] In this paper, we document the causal effects of social media exposure on hate crimes and xenophobic attitudes in Russia, a country with more than 180 ethnic groups. Furthermore, we use survey experiments to provide evidence of the particular mechanisms behind these effects.

Conceptually, social media may affect expression of hate, and hate crimes in particular, through different channels. First, social media can facilitate coordination and collective action: for example, Enikolopov et al. (forthcoming) show that social media facilitated the coordination of political protests in Russia in 2011-2012. Coordination through social media may be particularly relevant for illegal and stigmatized activities, such as hate crimes, as social media make it easier to find like-minded people through online communities and groups and possibly to out oneself as someone having such views. Second, social media may influence individual opinions: previously tolerant individuals might become exposed to intolerant views, while intolerant individuals might end up in "echo chambers" (Sunstein, 2001, 2017; Settle, 2018) that might make their views even more extreme. Finally, beyond changing attitudes, social media can may also affect people's perceptions of the acceptability of expressing hate, and therefore their willingness to express hate, conditional on holding a certain view. Indeed, individuals might be exposed to different reference groups that might

[1] For example, according to the Center for the Study of Hate and Extremism, across eight major metropolitan areas in the U.S., the number of hate crimes increased by more than 20% in 2016, which is significantly larger in both absolute and relative terms than any year-to-year increase in these cities since 2010.

[2] See, for example, "How Everyday Social Media Users Become Real-World Extremists," *New York Times*, April 25, 2018.

shape their perceptions about how society thinks of a certain view.

The main challenge in identifying a causal effect of social media is that access and consumption of social media are not randomly assigned. To overcome this challenge, we follow the approach of Enikolopov et al. (forthcoming) and exploit the history of the main Russian social media platform, *VKontakte* (VK). This online social network, which is analogous to *Facebook* in functionality and design, was the first mover in the Russian market and secured its dominant position with user share of over 90 percent by 2011. VK was launched in October 2006 by Pavel Durov, who was at the time an undergraduate student at Saint Petersburg State University (SPbSU). Initially, users could only join the platform by invitation through a student forum of the University, which had also been created by Durov. As a result, the vast majority of early users of VK were Durov's fellow students of SPbSU. This, in turn, made friends and relatives of these students more likely to open an account early on. Since SPbSU attracted students from across the country, this sped up propagation of VK in the cities these students had come from. As a result, the idiosyncratic variation in the distribution of the home cities of Durov's classmates had a long-lasting effect on VK penetration. This allows us to use fluctuations in the distribution of SPbSU students across cities as an instrument for the city-level penetration of VK. We then evaluate the effect of higher VK penetration on hate crimes and attitudes towards migrants using data on hate crimes collected between 2007 and 2015 by an independent Russian NGO, *SOVA*, as well as newly collected survey data on hate attitudes.

Using the instrumental variables approach, we show that penetration of social media led to more ethnic hate crimes, and that this effect is stronger in cities with a higher baseline level of nationalist sentiment prior to the introduction of social media. To proxy for baseline local nationalist sentiment, we use the city-level vote share of *Rodina* ("Motherland"), an explicitly nationalist and xenophobic party, in the 2003 parliamentary election, the last one before the creation of VK. We show that the impact of social media on hate crime victims

positively and significantly depends on the strength of pre-existing support of nationalists in the city: for example, a 10% increase in VK penetration increased hate crimes by 25.8% in cities where Rodina received most votes, but had zero effect in cities where Rodina got minimal support.

This stark heterogeneity is consistent with results on traditional media, which suggest that the impact of media on active manifestation of xenophobic attitudes depends on predispositions of the population. For example, Adena et al. (2015) demonstrate that radio propaganda by the Nazis in the 1930s was effective only in areas with historically high level of antisemitism, while Yanagizawa-Drott (2014), finds that social interactions allow the effect of traditional media (radio) on conflict to propagate. We further show that the effect of social media is stronger for crimes committed by multiple perpetrators (as opposed to those committed by single persons), consistent with social media likely playing a coordinating role.

To provide further evidence on the mechanisms behind the effect we next turn to the impact of social media on xenophobic attitudes of the population. To study these attitudes, we designed and conducted an online survey in the summer of 2018, with over 4,000 respondents from 125 cities.[3] The survey was positioned as a study of patterns of usage of social media and the Internet, to which we added the questions of interest that were related to ethnic hostility.

Given the potential for a stigma associated with directly reporting xenophobic views in a survey, we use the list experiment technique, one of the main methods to elicit truthful answers to sensitive survey questions (Blair and Imai (2012), Glynn (2013)) which was shown to perform particularly well in online surveys (Coutts and Jann (2011)).[4]. This approach gives our main measure of ethnic hostility, "elicited ethnic hostility."

[3]The survey and its analysis was pre-registered in the AEA RCT Registry (AEARCTR-0003066).

[4]The intuition behind this technique is that the respondents are asked only to indicate the number of statements with which they agree from a list. By adding the statement of interest to a random subgroup of respondents one can estimate the share of respondents agreeing with this statement without being able to identify who exactly agrees with it. See subsection 3.1.1 for more detail.

We also use this setup to infer whether social media could have affected this described stigma of reporting ethnic hostility. As mentioned before, it is conceivable that social media may affect perceptions of about the social acceptability of xenophobia. In the survey, We can measure this potential change in perceptions in a specific situation: communication in a survey. This admittedly does not capture the full extent to which the "change in stigma channel" might operate (it could be differentially relevant in other types of interactions), but might approximate what happens in a communication with strangers. To that end we use randomly included a direct question regarding negative attitudes toward other ethnicities, which we call "reported ethnic hostility."

Use the same IV approach we find a positive effect of social media penetration on elicited ethnic hostility, i.e. the share of respondents that hold xenophobic attitudes, regardless of whether they are willing to openly report them. The magnitude of the effect is particularly large in certain subsamples, specifically younger respondents and those with lower levels of education, i.e., groups more likely to use social media and to be engaged in hate crime.[5] Numerically, a 10% increase in VK penetration makes respondents 2.0% more likely to agree with the hateful statement in the list experiment.

We also investigate the effect of social media on self-reported ethnic hostility, i.e., the share of respondents who admit having xenophobic attitudes in a survey. In this case, we do not find a positive effect of social media on self-reported xenophobic preferences; if anything, the coefficients are negative, but generally insignificant. We obtain similar results if instead of our sample, we use the answers to the same direct question from a much larger, nationally representative survey of more than 30,000 respondents conducted in 2011 by one of the biggest Russian survey company, FOM (*Fond Obschestvennogo Mneniya*, Public Opinion Foundation).

[5]This goes in line with the argument in Boxell et al. (2017); Allcott and Gentzkow (2017) that the presumed impact of social media should be higher for groups more likely to be affected.

The difference between elicited and reported ethnic hostility provides a measure of the perceived stigma associated with the expression of such attitudes in a survey. Our results thus indicate that there is no evidence that social media reduced that perceived stigma. On the contrary, we find that, if anything, the perceived stigma increased as a result of social media exposure. This result also highlights the importance of using survey methods that reduce concerns of social acceptability bias, such as the list experiment: without these methods, we would be bound to erroneously find a negative or null effect of social media on xenophobic attitudes.

Finally, we show that our different results are all consistent with a simple model that captures the idea that social media increase the propensity of individuals to meet like-minded people, thereby resulting in higher polarization of opinions. We assume that each individual has a certain position, such as their attitude towards immigrants, but this position may change as a result of interaction with other people. To prevent convergence and ensure a nontrivial stationary distribution, we assume that each individual's attitude is subject to a random shock in every period.[6] If social media increase the propensity of individuals to meet like-minded people, this results in a society with a higher polarization of opinions, but with the same mean. The share of individuals who dislike immigrants beyond any given threshold (be it agreeing with the statement we provided or committing a hate crime) should therefore increase (and we argue below that the heterogeneity is as expected). At the same time, one should not expect a lower stigma of answering a direct question: indeed, higher polarization

[6]We thus employ a variant of DeGroot (1974) type of learning model, except that we assume that individuals adjust their political preferences rather than update their beliefs as a result of interactions with others. The model would be similar if we assumed that individuals learn about the optimal policy, such as the number of migrants that need to be admitted in the country. Like our paper, Dasaratha et al. (2019) introduces periodic shocks that may result in a nontrivial distribution in the long run. There, it is the object of social learning that is subject to shocks, and they show conditions under which opinions do not converge in a Bayesian framework. Here, we assume that individuals' opinions rather than the object of study are subject to shocks. These shocks may be interpreted as influence of books or news that individuals read, but the exact interpretation is not important; we merely aim at capturing some generic opinion formation process. See Golub and Sadler (2016) for an overview of models of learning in networks.

implies that the share of individuals who like immigrants also goes up, and social stigma of expressing xenophobia may go up as well, with the effect on the share of people answering the question affirmatively being ambiguous. The model therefore captures our results very closely, and while our empirical exercise was not designed as, and therefore is not, a proper test of it, we believe that this close connection between the theory and the empirics would stimulate further research in this area.

Our paper contributes to the growing literature on the impact of social media on polarization, xenophobia, and hate crime. Allcott et al. (2019), Mosquera et al. (2018), and Yanagizawa-Drott et al. (2019) provide evidence that social media makes people's political opinions more diverging. In contrast to these papers, we study more extreme outcomes, such as hate crime and hate attitudes. Qin et al. (2017) find that publications in the Chinese microblogging platform Sina Weibo predict future protests, strikes, conflicts, while Qin et al. (2019) show that the spread of information over online social networks leads to the spread of offline protests and strikes in China. Müller and Schwarz (2018) look at the relationship between social media and hate crime in Germany. Differently from our work, the paper focuses on short-run effects of social media during the week a particular content is posted, rather than the long-run effects of media penetration. Müller and Schwarz (2019) find that anti-Muslim hate crimes in the United States have increased in counties with high Twitter penetration users, but only since the start of Donald Trump's presidential campaign, and also analyze the effect of Trump's tweets on that type of hate crime. These fidings imply that social media can be instrumental for spreading incendiary messages from an important influencer, such as the president of the country. In contrast, our paper examines the long-run effect of penetration of social media on both hate crimes and hate attitudes, treating the content as endogenously formed. Moreover, we contribute to the literature by examining the underlying mechanisms behind the results, both empirically and theoretically.

This paper also contributes to a larger literature on the effect of media and, in particular,

social media on individual behavior. Enikolopov et al. (forthcoming), using an identification approach similar to ours, show that higher social media penetration increased the probability of political protests in Russia in 2011. In a similar vein, Manacorda and Tesei (2016) show that 3G penetration in Africa is associated with stronger cell-level protest participation. Bond et al. (2012) show that that political mobilization messages in Facebook increased turnout in the U.S. elections, Enikolopov et al. (2018) show that anti-corruption blog posts by a popular Russian civic activist had a negative impact on market returns of targeted companies and led to a subsequent improvement in corporate governance. Acemoglu et al. (2018) find that the protest-related activity on Twitter preceded the actual protest activity on Tahrir Square in Egypt. Steinert-Threlkeld et al. (2015) show that the content of Twitter messages was associated with subsequent protests in the Middle East and North Africa countries during the Arab Spring.

We also add to a growing literature studying the recent rise in populism and nationalist attitudes. Bursztyn et al. (2019) and Enke (2019) study the 2016 U.S. election. Algan et al. (2017) show that Great Recession triggered a trust crisis and led to higher voting shares of non-mainstream, particularly populist parties. Guriev et al. (2019) show that 3G penetration around the globe promoted populist voting and reduced government support. By also examining the effect of social media on the social acceptability of expressing intolerant views, this paper also relates to a growing literature that studies the role of social image concerns in a variety of settings (see DellaVigna et al. (2012) on charitable giving, DellaVigna et al. (2017) on voting decisions, Perez-Truglia and Cruces (2017) on campaign contributions, Bursztyn and Jensen (2015) on classroom participation, Bursztyn et al. (2018) on status goods, and Enikolopov et al. (2017) on political protests).

The remainder of this paper proceeds as follows. We discuss our identification strategy, data, and results on hate crimes in Section 2. In Section 3, we discuss survey results on xenophobic attitudes. We then present a model that reconciles our results from a unified per-

spective in Section 4. Section 5 concludes. The paper also includes three not-for-publication Appendices, with Appendix A containing all the proofs from Section 4, Appendix B containing additional tables, and Appendix C containing the survey script (translated into English).

2 Social Media and Hate Crimes

2.1 Identification Strategy

Our empirical strategy for identification of the causal effect of social media penetration follows the approach in Enikolopov et al. (forthcoming). In particular, we look at the penetration of the most popular social network in Russia, *VKontakte* (VK), which had substantially more users than than Facebook throughout the whole period we analyze. For example, in 2011, VK had 55 million users in Russia, while Facebook had 6 million users. VK was created in the fall of 2006 by Pavel Durov who at the time was a student at the Saint Petersburg State University (SPbSU). The first users of the network were largely students who studied with Durov at SPbSU. This made their friends and relatives at home more likely to open an account, which let to a faster spread of VK in these cities. Network externalities magnified these effects and, as a result, the distribution of the home cities of Durov's classmates had a long-lasting effect on VK penetration. In particular, the distribution of home cities of the students who studied at SPbSU at the same time as Durov predicts the penetration of VK across cities. This prediction is robust to controlling for the distribution of the home cities of the students who studied at SPbSU several years earlier or later. This effect persists throughout the period between 2007 and 2016 which we analyze, although the magnitude of the effect decreases over time. Thus, the effect of social media penetration is identified using a cross-sectional variation in the number of VK users across Russian cities, driven by the number of students from different cities who happened to study at SPbSU at the

time the network was created. The results of the first stage regression, similar to the one in Enikolopov et al. (forthcoming), are reported in Table A2 in the Online Appendix.[7][8] However, for the outcomes observed in the late 2010s, the first stage becomes weaker over time. As a result, for most of our empirical tests, the strength of the instruments is not always enough to make inference using conventional methods. Throughout the paper, we follow the recommendation in Andrews et al. (2019) and use the appropriate methods applicable in our particular case. In particular, in all tables we report weak instrument robust confidence sets developed by Chaudhuri and Zivot (2011) and Andrews (2017) and implemented in Stata by Sun (2018). Likewise, in all tables we denote the significance level of the endogenous coefficients based on these weak instrument robust sets and tests.

2.2 Data

The data on social media penetration and socioeconomic controls comes from Enikolopov et al. (forthcoming). The sample consists of 625 Russian cities with a population over 20,000 according to the 2010 Census.[9] To measure social media penetration we use information on the number of users of the most popular social media in Russia, VK. In particular, we calculate the number of VK users who report a particular city as their city of residence as of the summer of 2011. We use information on the city of origin of the students who studied at SPbSU based on the information provided in public accounts of the users of another social network, *Odnoklassniki* (Classmates). Specifically, we calculate the number of students coming from each city in five-year cohorts. We mostly focus on three cohorts in

[7]We use a more succinct set of controls than Enikolopov et al. (forthcoming), because we have a much smaller number of cities and we are facing power issues in survey analysis. The results of the analysis of the effect on hate crime are quantitatively and statistically similar to using exactly the same list of controls as in Enikolopov et al. (forthcoming).

[8]We also show that future VK penetration does not predict past nationalist party support in column (2) aof Table A2.

[9]The exceptions are cities with similar names that caused problems with disambiguation in the data, as well as Moscow and Saint Petersburg, which are excluded from the sample as outliers.

our analysis: i) those who were born the same year as the VK founder or within two years of it; ii) those who were born from three to seven years earlier than the VK founder; iii) those who were born from three to seven years later than the VK founder.

Data on hate crimes comes from the database compiled by SOVA Center for Information and Analysis.[10] This is a Moscow-based Russian independent nonprofit organization providing information related to hate crimes, which is generally considered to be the most reliable source of information on that issue. The dataset covers incidents of hate crimes and violent acts of vandalism, as well as convictions on any article of the Criminal Code relating to "extremism." These data are collected consistently starting 2007, with some incomplete data for 2004-2006. In the analysis we use data for 2007-2015. We classify all hate crimes as "ethnic" or "non-ethnic" based on the type of victim reported in the database. Table 1 presents more detailed information on the number of victims for each type. Based on the textual description of each incident in the database we have also manually coded the number of perpetrators for all the incidents. Non-ethnic crimes are more likely to be conducted by single perpetrators (see Figure 1), whereas ethnic hate crimes are more likely to be conducted by groups, with the modal number of perpetrators being two.

A potential concern with this data is that there could be a differential likelihoods of recording crimes across cities related to social media penetration, which could explain our results. Although We do not have evidence directly ruling out differential likelihoods of recording crimes, we believe that is highly unlikely that ethnic hate crimes were disproportionately more reported in areas with both higher penetration of VK *and* a higher baseline level of nationalist sentiment, *and* especially so for crimes with multiple perpetrators. We also provide evidence that the effects are stronger in larger cities, in which the likelihood of recording crimes being related to social media penetration is lower. Furthermore, our results on attitude changes are also consistent with social media having an effect beyond just the

[10]The database can be found at https://www.sova-center.ru/en/database/

reporting of hate crimes.

As a measure of nationalist sentiments in a city before the creation of the VK social network we use the vote share of the *Rodina* ("Motherland") party in the parliamentary election of December 2003, the only election this party participated in and the last parliamentary election before the creation of VK. This party ran on an openly nationalist platform. It received 9.2 percent of the vote and got 37 of the 450 seats in the State Duma, the lower house of the Federal Assembly of Russia. The data on electoral outcomes come from the Central Election Commission of the Russian Federation. We validate that the vote share for the party can serve as a proxy for nationalist sentiments by showing that it is positively and significantly correlated with ethnic hate crime in the subsequent years, as well as with xenophobic attitudes revealed in the opinion polls.

City-level data on population, age, education, and ethnic composition come from the Russian Censuses of 2002 and 2010. Data on average wages come from the municipal statistics of RosStat, the Russian Statistical Agency. Additional city characteristics, such as latitude, longitude, year of city foundation, and the location of administrative centers, come from the Big Russian Encyclopedia.[11]

The data on attitudes towards other ethnicities come from a survey of over 4,000 individuals that we conducted in the summer of 2018 in 125 Russian cities. The survey was conducted by a professional marketing firm, *Tiburon Research*, with a representative panel of urban Internet users in Russia. The sample consists of 4,327 respondents, of which 2,166 were allocated to the control group and 2,161 to the treatment group.[12]

[11]The electronic version of the Encyclopedia can be found at https://bigenc.ru/

[12]We collected the data in two batches, the pilot and the main experiment. As part of the pilot, we surveyed 1,007 individuals from 20 cities. Individuals from this batch were randomized into three groups, with one containing a statement about ethnic minorities as part of the list experiment, another containing a statement about LGBTQ individuals, as well as a control group. As we found no reliable data on hate crimes against LGBTQ individuals, we dropped the second group of 336, leaving us with 671 individuals from the pilot. As part of the main experiment, we surveyed 4,034 individuals from 111 cities. In this batch, the cities were randomly chosen by the firm we were working with, and since we had the data on VK penetration for only 105 of these cities, we had to drop 246 observations from six cities. Additional 12 surveys were

We also use data from the MegaFOM opinion poll conducted by FOM (*Fond Obschestvennogo Mneniya*, Public Opinion Foundation) in February 2011. This is a regionally representative survey of 54,388 respondents in 79 regions of Russia, of which 29,780 respondents come from 519 cities in our sample. In particular, we use information on answers to exactly the same direct question about hostility to different ethnicities that was asked in our survey conducted in 2018.

2.3 Social Media and Hate Crime: Empirical Specification

Our main hypothesis is that social media penetration (specifically, VK penetration) has an impact on hate crime. Thus, we estimate the following model:

$$\text{HateCrime}_i = \beta_0 + \beta_1 \text{VKpenetration}_i + \beta_2 \mathbf{X_i} + \varepsilon_\mathbf{i}, \tag{1}$$

where $HateCrime_i$ is a measure of hate crime, which reflects either the total number of victims of hate crimes in city i during the period 2007-2015, or the number of victims of particular types of hate crime (ethnic or non-ethnic crimes, conducted by single or multiple perpetrators). We also consider three sub-periods 2007-2009, 2010-2012, and 2013-2015 separately. VKpenetration_i is the logarithm of the number of VK users in city i in summer 2011.[13] This endogenous variable is instrumented using the number of students from each city in a five-year student cohort who have studied at the same year as the founder of VK, Durov, as well as one or two years earlier or later. $\mathbf{X_i}$ is a vector of control variables that include the number of students from the city in the other two five-year student cohorts, those that studied three to seven years earlier than Durov, and those that studied three

incomplete, which left us with 3,776 observations from the main part. In most analyses, we pool the two batches together, but our results are robust to looking at the second batch only. The survey was approved by the University of Chicago Institutional Review Board (IRB18-0858) and was pre-registered in the AEA RCT Registry (AEARCTR-0003066).

[13]We add one to the variable in our logarithm measures to deal with zeros.

to seven years later than Durov. It also includes the following socioeconomic controls: the logarithm of the population, the indicator for being a regional or a subregional (*rayon*) administrative center, the average wage in the city, the number of city residents of different five-year age cohorts, the share of population with higher education in 2010 in each five-year age cohort, the indicator for the presence of a university in the city, ethnic fractionalization, and the logarithm of the number of Odnoklassniki users in 2014. For all specifications we report weak-instrument robust confidence sets.[14] Similarly, for our heterogeneity analysis we estimate the equation:

$$\text{HateCrime}_i = \beta_0 + \beta_1 \text{VKpenetration}_i \times \text{Nationalist Support}_i + \beta_2 \mathbf{X_i} + \varepsilon_\mathbf{i}, \tag{2}$$

where $\text{NationalistSupport}_i$ denotes the votes for the nationalist Rodina party in 2003 and $\mathbf{X_i}$ is the new set of controls.

2.4 Social Media and Hate Crime: Results

Table 2 summarizes the results of estimating Equation (1) for the average impact of exposure to VK on hate crime. There is no consistent evidence of a significant effect of VK penetration on hate crime, for either ethnic- or non-ethnic- hate crime or for crimes conducted by single or multiple perpetrators. At the same time, the confidence intervals do not allow us to rule out large effects (e.g., at maximum 57 percent increase, i.e. 0.58 of a standard deviation of the dependent variable in column 1), though only one out of nine coefficients in the table is marginally significant.

However, this approach masks an important heterogeneity of the effect with respect to the underlying level of nationalism. People in cities with very few nationalists to begin with

[14] As discussed above, we report weak instrument robust confidence sets developed by Chaudhuri and Zivot (2011) and Andrews (2017) and implemented in Stata by Sun (2018) throughout the paper.

and people from very nationalist cities can respond differently to the arrival of social media. To capture this dimension of heterogeneity into account, we interact VK penetration with a measure of pre-existing nationalist support, as captured by the Rodina party vote share in 2003.

Table 3 summarizes the results of estimating Equation (2). The nationalist party support variable is demeaned to simplify interpretation of the direct coefficient. In all specifications except one the effect of social media penetration on hate crime is significantly stronger in cities with higher preexisting level of nationalism. Numerically, the results imply that the effect of a 10% increase in social media penetration ranges from being close to zero (non-significant with different signs) at the minimum level of nationalist party support to a 25.8% increase in total number of hate crime victims at the maximum level of nationalist support (column 1 of Table 3).

The results indicate that in cities with high pre-existing level of nationalism, social media increased the total number of victims of hate crimes. This is true for the victims of ethnic and non-ethnic crimes, as well as of crimes conducted by either single or multiple perpetrators. In other words, social media spurs acts of hate crime in places with higher levels of pre-existing nationalism. Another important takeaway from Table 3 is that the coefficient of interest is noticeably larger for incidents that involved multiple perpetrators, i.e., acts of violence that require coordination.[15] At the same time, the results are significant for crimes with single perpetrators as well (with the exception of non-ethnic crime in column 8), which suggests that while social media facilitated coordination and thus contributed to hate crime, coordination alone does not fully explain the overall impact of social media.

To interpret the evidence on the link between social media and hate crime victims presented in Tables 2 and 3, it is important to distinguish between the intensive and extensive

[15]In the seemingly unrelated regressions specification the difference between the interaction coefficients in columns 2 (single perpetrator) and column 3 (multiple perpetrators) is statistically significant at the 10% level; the differences for ethnic and non-ethnic crimes are similarly large in magnitude.

margins. In Table A3, we estimate equation 2 with the number of crimes rather than the number of victims as the dependent variable. The results suggest that the number of crimes responds to the introduction of social media and to the number of victims very similarly, both in terms of magnitude and statistical significance. For example, the impact of 10% increase in social media of social media penetration on the number of crimes is bounded by 24.8% for total crimes, a figure very similar to the maximal effect on the number of victims. In other words, the increase in the number of hate crime victims is well explained by the increase in the number of crimes, so it is the extensive margin that seems to play the role.

We also attempt to understand the evolution of the impact of social media over time. The beginning of our time period, 2007-2009, was the time of a rapid introduction of social media into people's lives, with the total number of VK users growing from hundred thousand to more than thirty million users, while by 2013-2015 the exponential growth had already stopped and other platforms, such as Twitter, started to gain some popularity. At roughly the same time, following the Arab Spring and the protests in Russia in 2011-2012, the Russian government began to regulate online content, which prevented openly xenophobic communities from being created and sustained. If we examine the effect for the three 3-year sub-periods separately (see Table A4), one can see that the effects are similar in size in 2007-2009 and 2010-2012, but become noticeably smaller and statistically insignificant in 2013-2015. We should note, however, that the differences in coefficients for the later (2013-2015) and earlier (2007-2009, 2010-2012) periods is not statistically significant in a seemingly unrelated regressions framework. On top of that, the predictive power of the instrument in the first stage regression is going down with time (see Figure A1). Thus, while our findings are consistent with abatement of the impact of social media over time, we should interpret these intertemporal results with caution.

Table 4 reports the results of placebo regressions for hate crime in the period 2004-2006, i.e., before the creation of the VK social network. The results indicate no significant effect

of social media on hate crime even in cities with maximum level of support of the nationalist party, and the difference between these results and the results in Table 3 is statistically significant in seemingly unrelated regressions framework.[16] The null results in Table 4, however, may be driven by the fact that the data for this time period are incomplete, in contrast to the later years.

As was mentioned in Section 2.2 a potential consern is that the results are driven by differential likelihoods of recording crimes that is correlated with explanatory variables. Although We do not have direct evidence directly ruling out differential likelihoods of recording crimes, we can check if the effects that we identify depend on the size of the cities. Arguably, in smaller cities reporting of hater crimes may be more dependent on whether they were discussed in social media or not, which should make the measurement error stronger in smaller cities. However, similarly to the findings in (Enikolopov et al. (forthcoming)), if we restrict the sample to cities with population above the median, the results become only stronger (see Table A5 in the Online Appendix). In addition, it is highly unlikely that ethnic hate crimes were disproportionately more reported in areas with both higher penetration of VK *and* a higher baseline level of nationalist sentiment, *and* especially so for crimes with multiple perpetrators. We also our results on attitude changes are also consistent with social media having an effect beyond just the reporting of hate crimes (see the next section).

Overall, the results in Tables 2-4 indicate that social media had a positive effect on hate crime, but only in places where the level of nationalism was already sufficiently high before the creation of social media.

[16]There are not enough observations of non-ethnic crimes with single perpetrators for that period to estimate the results.

3 Social Media and Hate Attitudes

The results so far can be explained by various mechanisms in play. More specifically, social media can increase hate and hate crime through:

(1) coordination – it is easier to find like-minded people online and coordinate activities (offline meetings) that might eventually lead to hate crimes;

(2) persuasion – social media can change people's opinions and make previously tolerant people more intolerant toward minorities, while previously intolerant people could become even more intolerant;

(3) social acceptability – social media can make people more willing to express views that they previously were reluctant to express in public.

Numerical differences between crimes committed by single and multiple perpetrators (combined with the effects being driven by cities with stronger pre-existing nationalism) point out toward coordination being one of the explanations. However, as noted above, the results on crimes with a single perpetrator suggest that mechanisms other than coordination should be at play as well. To further explore the mechanisms behind the effect of social media on hate crime, we designed and conducted a survey aimed at measuring the true level of underlying nationalism expressed in an anonymous way through the use of a list experiment. Examining the effect of social media on implicit xenophobic attitudes will allow us to see if one of the mechanisms through which social media affects hate crime involves changing people's preferences and persuading them to become more nationalist.

As part of this survey, we also measured self-reported intolerance towards migrants as the share of respondents who admitted such attitude in response to a direct question. By examining how the effect differs for the self-reported xenophobia as compared to the xenophobia elicited through the list experiment, we are able to check if social media affected the social stigma of expressing xenophobic attitudes openly.

3.1 Survey Evidence

To measure implicit xenophobic attitudes we conducted an online survey in 2018, with a list experiment embedded as part of it. This design (also called the "unmatched count" and the "item count technique", originally formalized by Raghavarao and Federer (1979) and further developed, in recent works by Blair and Imai (2012) and Glynn (2013), among others) is a standard technique for eliciting truthful answers to sensitive survey questions. The list experiment works as follows. First, respondents are randomly assigned to either the control group or the treatment group. Subjects in both groups are then asked to indicate the number of statements they agree with. In this way, the subject never reveals their agreement with any particular statement (unless the subject agrees with all or none, which is something the experimental design should try to avoid), only the total number of statements. In the control condition, the list contains a set of statements or positions that are not stigmatized. In the treatment condition, the list includes all the statements from control list, but also adds the statement of interest, which is potentially stigmatized (and in both cases, the positions of statements are randomly rotated). The support for the stigmatized opinion can then be inferred by comparing the average number of statements the subjects agree with in the treatment and control conditions. For recent applications of list experiments in economics, see Enikolopov et al. (2017) and Cantoni et al. (2019).

In our case, the survey participants were asked the following question: *"Consider, please, whether you agree with the following statements. Without specifying exactly which ones you agree with, indicate just the number of statements that you can agree with."* The respondents in the control group were given four statements unrelated to the issues of ethnicity.[17] The respondents in the treatment group were given the additional fifth statement: *"I feel*

[17]The exact statements were the following: i) Over the week I usually read at least one newspaper or magazine; ii) I want to see Russia as a country with a high standard of living; iii) I know the name of the Chairman of the Constitutional Court of the Russian Federation; iv) Our country has a fairly high level of retirement benefits.

annoyance or dislike toward some ethnicities." Here, we took the exact wording used by one of the leading opinion polling firms in Russia in their regular large scale surveys, which has the additional advantage of making our results comparable with the results of the opinion polls by this firm (see subsection 3.1.5 for more detail). Respondents in the control group, after answering the question on the number of statements they agreed with (which did not include the statement on ethnicities), were then asked about annoyance or dislike toward some ethnicities directly. Overall, the share of respondents who agreed with the xenophobic statement in the list experiment (i.e., the difference between the average number of statements with which respondents in treatment and control group agreed) was approximately 38 percent, while the percentage of respondents who admitted being xenophobic in the direct question was 33 percent.

3.1.1 Elicited hostility, individual-level results

Given the randomization, comparing the mean number of positive answers between treatment and control groups provides a valid estimate of the percentage of respondents who agree with the sensitive statement about having xenophobic attitudes (Imai (2011)). However, our goal is to estimate the impact of an independent variable (social media penetration) on the answer to this sensitive question. Following Imai (2011) and Blair and Imai (2012), we use the regression model with interactions to estimate how answers to the list experiment question depend on other parameters, in our case characteristics of the respondent's city. Formally, we estimate the following model:

$$\text{NumberOfStatements}_{ij} = \beta_0 + \beta_1 \text{T}_{ij} + \beta_2 \left(\text{T}_{ij} \times \text{VK}_j\right) + \beta_3 \text{VK}_j + \beta_4 \text{X}_{ij} + \varepsilon_{ij}, \qquad (3)$$

where $\text{NumberOfStatements}_{ij}$ is the number of statements with which respondent i from city j agreed, T_{ij} is the dummy variable for whether respondent i from city j was assigned to the

treatment group, and $textnormal{VK}_j$ is the measure of VK (social media) penetration in city j instrumented by the number of students from the city who studied at SPbSU together with the founder of VK, controlling for the number of students from older or younger cohorts. Other controls include city level controls and the interaction of pre-existing nationalism with the treatment dummy to account for the differential response. Standard errors are clustered at the city level.

In this specification, the effect of social media penetration on the share of respondents in city j who have implicitly xenophobic attitudes is captured by coefficient β_2. In what follows, we also look at the subsamples, paying special attention to the groups more likely to be involved in hate crime (males, younger respondents (below the median age in the sample, which is 32), and respondents with lower level of education (below the median in our sample)).[18]

The results presented in Table 5 indicate that social media increases elicited hostility to other nationalities, both on average and for subgroups of population that are more likely to engage in hate crime (male, young, or low-educated). The results in column 1 imply that, on average, the elasticity of elicited hostility with respect to social media penetration is 0.075. In other words, a 10% increase in VK penetration increases the share of those agreeing with the statement in the list experiment by 4.5%.[19] This magnitude goes up to 6.8% for males (column 2), 9.0% for those with low education (column 4), and 5.3% for younger respondents (column 6). We did not find any significant effect of VK for females, those with higher education, or older respondents, and the magnitude of coefficients is considerably smaller for these groups than their opposites. We should note that this whole setup is essentially

[18]Note that we pre-registered heterogeneity by gender in our pre-analysis plan, but later we decided that these other simple characteristics (being young and low-educated) are also likely to predict being a hate crime perpetrator, thus we added them to the analysis.

[19]We got this number by dividing one tenth of the effect, 0.0083, by the baseline level of those agreeing with the xenophobic statement in the absence of VK, as given by the direct coefficient for the list experiment option in the whole sample, 0.185. For the other columns, similar calculations apply.

an intention-to-treat framework, as not all survey respondents or their friends actually use VK, and we do not have an instrument for the exposure to VK at the individual level. As a result, the numbers in Table 5 may be interpreted as a lower bound for the true effect.

3.1.2 Elicited hostility, city-level results

In this subsection, we repeat the analysis above at the city level.[20] Let us denote the variable NumberOfStatements$_{ij}$ as y_{ij}. Then, assuming that Equation (3) is a true data generating process, we derive the city-level specification we would like to estimate. More specifically, we first sum individual responses by city and treatment status:

$$\sum_{T_{ij}=0} y_{ij} = \beta_0 \sum_{T_{ij}=0} 1 + \beta_3 \mathrm{VK}_j \sum_{T_{ij}=0} 1 + \beta_4 \sum_{T_{ij}=0} \mathrm{X}_{ij} + \sum_{T_{ij}=0} \varepsilon_{ij};$$

$$\sum_{T_{ij}=1} y_{ij} = (\beta_0 + \beta_1) \sum_{T_{ij}=1} 1 + (\beta_2 + \beta_3)\, \mathrm{VK}_j \sum_{T_{ij}=1} 1 + \beta_4 \sum_{T_{ij}=1} \mathrm{X}_{ij} + \sum_{T_{ij}=1} \varepsilon_{ij}.$$

We then divide both sides of the last two equations by the number of respondents in each treatment group in a city ($\sum_{T_{ij}=a} 1$) and take the difference. We get

$$\frac{\sum_{T_{ij}=1} y_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} y_{ij}}{\sum_{T_{ij}=0} 1} = \beta_1 + \beta_2 \mathrm{VK}_j + \beta_4 \left[\frac{\sum_{T_{ij}=1} X_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} X_{ij}}{\sum_{T_{ij}=0} 1} \right] + \eta_j, \qquad (4)$$

here we denoted $\left[\frac{\sum_{T_{ij}=1} \varepsilon_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} \varepsilon_{ij}}{\sum_{T_{ij}=0} 1} \right]$ as η_j to simplify notation.

All city-level controls that were not interacted with an extra treatment option T_{ij} cancel each other in (4). For a conservative estimation without simple demographic controls, the only term that was interacted and that differs between treatment and control group is

[20] This is the main specification mentioned in our pre-registration.

$\text{NationalistSupport}_j \times T_{ij}$. Thus, the city level specification reduces to

$$\frac{\sum_{T_{ij}=1} y_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} y_{ij}}{\sum_{T_{ij}=0} 1} = \beta_1 + \beta_2 \text{VK}_j + \beta_{4,ns} \text{NationalistSupport}_j + \eta_j. \tag{5}$$

We present the results of this estimation in Table 6. As one can see, the results are largely consistent with the results at the individual level (Table 5), though the coefficients in Table 6 are slightly larger in terms of magnitudes.

Overall, the results in Tables 5 and 6 indicate that social media penetration had a positive effect on the share of people who have implicit xenophobic attitudes, and more so among the groups of respondents likely to be involved in hate crimes (and, in the case of younger and low-educated individuals, groups that are arguably likely to be persuadable). These findings speak in favor of the persuasive effect of social media on xenophobic attitudes.

3.1.3 Self-reported hostility and stigma

The effect of social media on self-reported xenophobic attitudes is estimated at the individual level using the following specification:

$$\text{SelfReportedHate}_{ij} = \beta_0 + \beta_1 \text{VK}_j + \left[\beta_2 \text{ElicitedHostility}_j\right] + \beta_3 \text{X}_{ij} + \varepsilon_{ij} \tag{6}$$

Here $\text{Elicited Hostility}_j$ is the average difference between the numbers of statements that participants from the treatment group and the control group in city agreed with, i.e., $\frac{\sum_{T_{ij}=1} y_{ij}}{\sum_{T_{ij}=1} 1} - \frac{\sum_{T_{ij}=0} y_{ij}}{\sum_{T_{ij}=0} 1}$ from city-level equation (5).

The results without controlling for the results of the list experiment are reported in Table 7, Panel A. The coefficient of interest, VK_j, is generally not statistically significant and has a negative sign. For one particular specification in which we look at the subset of younger respondents (column 4), 95% weak-instrument-robust confidence set lies entirely below zero. Thus, we find no evidence that social media reduces stigma associated with

expression of hateful opinions. Even though social media seems to increase actual hostility to other ethnicities (Tables 5 and 6), it does not decrease (and if anything, increases) the stigma associated with expressing xenophobic attitudes in public.

In Table 7, panel B we report the effect of social media on self-reported intolerance when the elicited level of hostility is controlled for. Unfortunately, here we hit the limits of our identification approach, with weak instrument robust confidence sets being very imprecise and some of them even including the entire grid. However, the results for the city-level estimation of (6) are qualitatively similar and are presented in the Appendix in Table A6.

Overall, our survey analysis implies that in cities with higher social media penetration respondents are more likely to have implicit xenophobic attitudes, but at the same time are not more likely to express them openly to a stranger, such as surveyor (of course, we cannot rule out differential changes in perceived social acceptability vis-a-vis other audiences, such as neighbors). In Section 4 below we offer a theoretical model that shows that both of these findings are consistent with an increased polarization caused my social media.

3.1.4 Interaction with pre-existing nationalism

In Section 2 we showed that pre-existing nationalism increases the effect of social media on hate crimes. It is natural to wonder whether the effect of social media on hate as measured by the survey is similarly affected. Unfortunately, the variation in the survey data is not sufficient to identify the interaction term with a reasonable precision. The instrument turns out to be too weak for a meaningful analysis (see Table A7 in the Online Appendix) and weak-instrument-robust confidence sets for this estimation include the entire grid. The results based on city-level data are presented in Table A8, but they should be interpreted with extreme caution for two reasons. First, they are based on a small number of observations per city and the number of observations varies significantly from city to city. Second, and, most importantly, to the best of our knowledge there is no standard way of computing standard

errors in this case, which involves both aggregation of noisy data and weak instruments.

With these caveats, the results in Table A8 suggest that the interaction term between VK penetration and pre-existing nationalism is *negative* in the whole sample, as well as for male, low-educated, and young subsamples. In other words, the increase in elicited hate is *smaller* in cities where pre-existing nationalism was higher. This is particularly interesting and perhaps surprising in light of our results on hate crimes in Table 3, where the corresponding interaction term is *positive*. Nevertheless, our model, which we present below, reconciles and explains both of these results.

The interaction results for the self-reported hate are presented in Table A9. As in Tables 7 and A6, the estimates are too noisy to draw any conclusions.

3.1.5 Results from the 2011 survey

To make sure that the lack of an effect on self-reported xenophobic attitudes is not a consequence of the timing of the survey (almost twelve years after VK was founded) or the number of respondents, we replicate the analysis of our own survey using data from a much larger survey conducted in February 2011. This MegaFOM opinion poll, conducted by FOM (*Fond Obschestvennogo Mneniya*, Public Opinion Foundation), has a regionally representative sample of 54,388 respondents in 79 regions, of which 29,780 respondents come from the 519 cities in our sample. This survey contained a direct question on dislike toward other ethnicities with exactly the same wording as the question we used to measure self-reported hostility in our survey, which we analyzed above.

The results of estimating equation (6) based on this sample are presented in Table 8. These results indicate that, as in the case of the 2018 survey, there is no significant relation between social media penetration and self-reported xenophobic attitudes. This null result holds regardless of the initial level of nationalism in a city. Weak instrument robust confidence sets are, again, too large to claim that these are indeed zero results, though for the

direct effect we can rule out a more than 25% increase in reported xenophobic attitudes following a 10% increase in social media penetration.

4 Model

We now present a simple model of social learning to show that our empirical findings can be explained by social networks increasing individuals' propensity to meet like-minded people.

4.1 Social networks and distribution of preferences

Time is discrete and infinite, $t = 0, 1, 2, \ldots$, and there is a continuum of individuals in a society. Each individual has a political position over some dimension of interest, such as xenophobia. This political position may be interpreted as taste-based (e.g., whether the individual likes or hates immigrants) or an opinion about a particular policy (e.g., the number of immigrants to be allowed, or the minimal requirements such as education and lack of criminal history that they must satisfy). Importantly, an individual's political position may evolve over time, so we write x_i^t to denote the position that individual i has at time t. The positions at time 0, x_i^0, are taken exogenously from some distribution H^0 with c.d.f. F^0 with finite first (denoted by $\mu^0 = \mathbb{E}x_i^0$) and second moments; we assume for simplicity that there is a continuum of individuals at each political position.

In each period starting from $t = 1$, individuals may change their political position. Assume that their new position will incorporate their current one with weight ω, and the positions of other people they talk to with weight $1 - \omega$. In each period, they talk to a continuum of other people, share τ of which are just like them (i.e., with the same political position), and share $1 - \tau$ are random individuals from the society. To capture the idea that social networks make it easier for like-minded individuals to find each other and spend time with them, we condsider τ to be a proxy for penetration of social networks. In addition, we

assume that each individual's position is subject to a random additive shock ε_i^t, which has normal distribution $\mathcal{N}(0, \sigma_\varepsilon^2)$; these shocks are independent across individuals and time.[21] We thus have the following evolution of opinion of individual i:

$$x_i^t = (\omega + (1-\omega)\tau) x_i^{t-1} + (1-\omega)(1-\tau)\mathbb{E}x_{-i}^{t-1} + \varepsilon_i^t, \tag{7}$$

where $\mathbb{E}x_{-i}^{t-1}$ equals the integral over political positions of other people in the society in the previous period.

Lemma 1. *The distributions of political positions in the society, $F^0, F^1, F^2, \ldots$, converge in distribution to $\mathcal{N}(\mu, \sigma^2)$ as $t \to \infty$, where $\mu = \mu^0 = \int_{-\infty}^{+\infty} x dF^0(x)$ is the mean of the initial distribution and σ^2 is given by*

$$\sigma^2 = \frac{\sigma_\varepsilon^2}{1 - (\omega + (1-\omega)\tau)^2}, \tag{8}$$

which is increasing in σ_ε^2, ω, and τ.

In other words, this model of social learning with shocks predicts convergence of the distribution of preferences to a normal one, with mean equal to the mean of the original distribution, whereas all the other information about the original distribution is lost over time. The variance of the limit distribution is nontrivial because of persistence shocks to preferences. The more individuals are influenced by people with random opinions, the faster these preference shocks dissipate, and the smaller the variance of the limit distribution is. Conversely, if people are mostly influenced by themselves (higher ω) or like-minded people (higher τ), as in 'echo chambers,' the limit distribution has a higher variance, so in other

[21] The shocks are best thought of as idiosyncratic, but it is easy to amend the model so that these shocks capture influence by sources that maintain their distribution over time. For example, these might come from general human knowledge (say, books that individual i might read in period t) or influence by a certain group of individuals (politicians, celebrities, religious leaders) who have fixed positions that do not evolve over time.

words, the society is more polarized.[22]

4.2 Extreme political preferences

We now study how support for different political positions is affected by increased penetration of social networks. By Lemma 1, an increase in τ results in a more polarized limit distribution, i.e. one with a higher variance σ. We therefore need to study the effects of an increase in σ.

Take any cutoff q and consider the shares of individuals with preferences to the left and to the right of q. Denote these shares by L_q and R_q, respectively, so

$$\begin{aligned} L_q &= \Pr\left(x_i^\infty < q\right) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{q} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \\ R_q &= \Pr\left(x_i^\infty > q\right) = \frac{1}{\sqrt{2\pi}\sigma}\int_{q}^{+\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx. \end{aligned}$$

It is straightforward to prove the following result (we formulate it for R_q only, as $L_q = 1 - R_q$):

Proposition 1. *Suppose that $\mu < q$. Then R_q is increasing in σ, so $\frac{\partial R_q}{\partial \sigma} > 0$. The magnitude of this effect is nonmonotone in μ: $\frac{\partial^2 R_q}{\partial \sigma \partial \mu} > 0$ for $\mu \in (-\infty, q-\sigma)$ and $\frac{\partial^2 R_q}{\partial \sigma \partial \mu} < 0$ for $\mu \in (q-\sigma, q)$. Similarly, if $\mu > q$, then $\frac{\partial R_q}{\partial \sigma} < 0$, $\frac{\partial^2 R_q}{\partial \sigma \partial \mu} < 0$ for $\mu \in (q, q+\sigma)$ and $\frac{\partial^2 R_q}{\partial \sigma \partial \mu} > 0$ for $\mu \in (q+\sigma, +\infty)$.*

In other words, the opinion of a relative minority (R_q if $\mu < q$ or L_q of $\mu > q$) becomes more popular as variance σ increases. However, the magnitude of the effect is the highest for values of q about one standard deviation from the mean μ, and it vanishes for values either very far from or very close to the median. For the former, the density is too low to have an

[22] Dasaratha et al. (2019) study a society of Bayesian individuals that learn about an ever-changing state of the world, in which case the shocks can correspond to new private signals that individuals get. Such a model would generate similar comparative statics; for simplicity we focus to DeGroot (1974) type of learning with shocks, as in (7).

effect, whereas for the latter, both probabilities are close to 0.5 and their difference is small. The minimum is attained at the inflection point of the bell curve, which is illustrated on Figure 1.

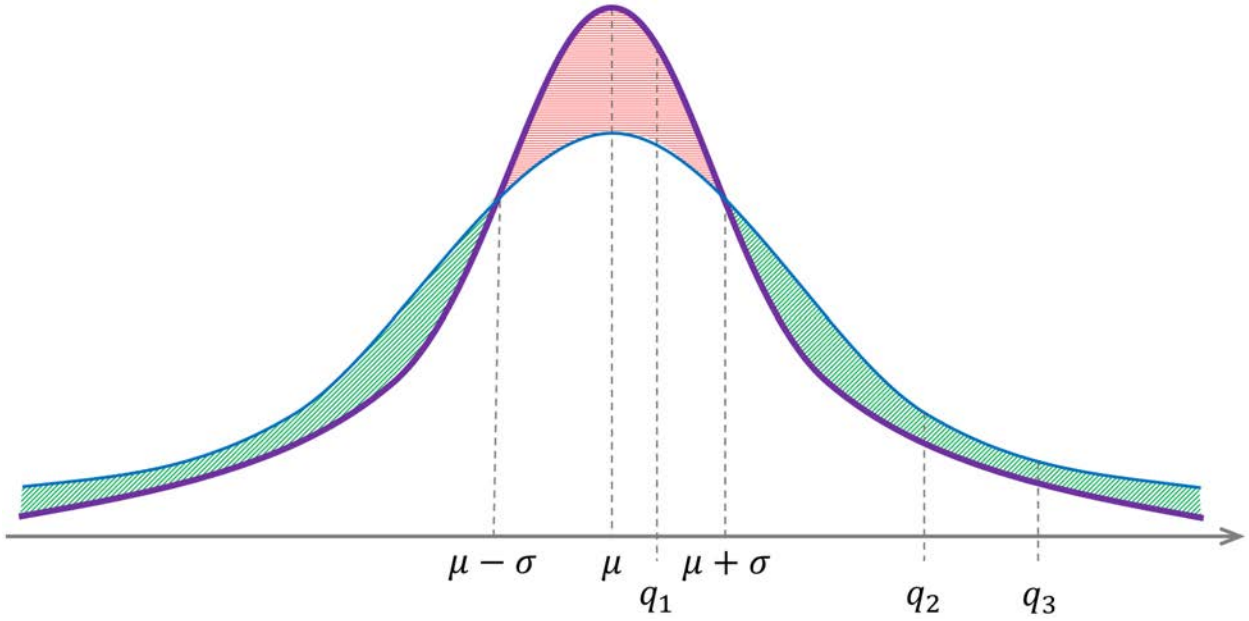

Figure 1: Distributions of political positions before (purple/thick) and after (blue/thin) penetration of social networks. For a given cutoff, the increase in the share of individuals with positions above this cutoff is the area between the curves to the right of the cutoff, with green/diagonally shaded area taken with a positive sign and the red/horizontally shaded area with a negative one.

This proposition has important implications for our setting. Suppose that the tolerance-xenophobia axis is oriented such that higher values correspond to stronger xenophobia. On this axis, there is some cutoff q_1 corresponding to the person sometimes experiencing antipathy towards other nationalities (the question we are asking in the list experiment). There is some cutoff q_2 corresponding to the person being just willing to commit a hate crime as part of a group, and some cutoff q_3 corresponding to him being just willing to commit a hate crime alone. It is natural to expect that $q_1 < q_2 < q_3$, but more importantly, all such xenophobic preferences are expressed by a minority of people. This means that all these cutoffs exceed μ, and so by Proposition 1 a higher σ, e.g. induced by the availability of social

media, should increase the share of people with xenophobia exceeding any of these cutoffs. In other words, more people should dislike migrants (consistent with the individual-level and city-level results of Section 3.1.1 and 3.1.2, see Tables 5 and 6), and more people should commit hate crimes, both individually and jointly (consistent with the results of Section 2.4, see Table 2), in places with higher social media penetration.

Let us now look at how these effects depend on preexisting nationalism, which is naturally captured by μ, with higher μ corresponding to more nationalism and xenophobia. It is reasonable to think that the cutoffs that guide whether a person commits a hate crime when given an opportunity, either individually (q_3) or jointly (q_2), lie more than a standard deviation above the median, so we should have $q > \mu + \sigma$. Indeed, for a normal distribution, the mass of distribution on the right of $\mu + \sigma$ equals $F(-1) \approx 0.16$, which is certainly higher than the number of potential perpetrators in our setting.[23] For these values, Proposition 1 implies that an increase in μ (or, equivalently, a decrease in q) would increase the derivative $\frac{\partial R_q}{\partial \sigma}$. In other words, a higher level of preexisting nationalism leads to a stronger effect of social networks on hate crime, consistent with the results of Section 2.4 (See Table 2). Conversely, for the cutoff that determines an affirmative answer to the statement we provided (q_1), we should have $\mu < q < \mu + \sigma$ in our setting, because the share of people agreeing with this statement is about 38%. For this range, Proposition 1 suggests that an increase in mu would have an opposite effect, decreasing the derivative $\frac{\partial R_q}{\partial \sigma}$. Thus, a higher level of preexisting nationalism would alleviate the effect of social networks on elicited hate. As discussed in Section 3.1.4, (Tables A5 and A6), we should interpret the corresponding empirical results with caution because of weak instruments, but the signs of the point estimates for the interaction terms are consistent with this prediction.

[23] Hate crime is still a relatively rare phenomenon in modern Russia, with the share of perpetrators well below 1% of the population in all the cities that we consider.

4.3 Self-reported support for extreme positions

Consider an individual i with position x_i who is asked, before an audience and therefore under social pressure, whether it exceeds q. Denote the affirmative answer by $d_i = Y$ and the negative answer by $d_i = N$. The individual gets disutility from expressing preferences that are far from his/her own, or to put it another way, there is a cost of lying. Specifically, if $x_i > q$ and s/he chooses $d_i = N$, s/he gets disutility $h(x_i - q)$, where $h(\cdot)$ is an increasing continuous function with $h(0) = 0$; in other words, we assume that egregious lies are more costly than little lies. Similarly, if $x_i < q$ and s/he chooses $d_i = Y$, s/he gets disutility $h(q - x_i)$. In both cases, telling the truth does not yield direct utility or disutility.

The individual also cares about social approval. We assume that i's response to the question whether x_i exceeds q is observed by a random other individual in the society (assuming that it is observed by several or even all individuals leads to a very similar model with similar results). This other individual j will form a posterior belief about the individual i's type. We assume that individual with political position x_j dislikes individual with position x_i according to a function $g(x_j - x_i)$. Thus, individual i chooses answer d_i to maximize his utility U_i that consists of (negative) direct cost C_i and social cost S_i:

$$\begin{aligned} & U_i(d_i, q) = -C_i(d_i, q) - S_i(d_i, q) \\ = \; & -\mathbf{I}_{\{x_i > q \wedge d_i = N\}} h(x_i - q) - \mathbf{I}_{\{x_i < q \wedge d_i = Y\}} h(q - x_i) \\ & - \int_{-\infty}^{\infty} \mathbb{E}_{-i}(g(x - y) \mid d(x) = d_i) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy. \end{aligned}$$

The latter term $S_i(d_i)$ captures the expectation of $g(x_i - y)$ by an observer with position y who knows that individual i chose action d_i, and then the expectation is taken over the possible realization of observer's types.

In general, the game admits multiple equilibria because of strategic complementarity

of adherence to social norms; however, if individuals are sufficiently averse to lying the equilibrium is unique. We will impose the following sufficient condition; if it does not hold, then the comparative statics result are true for equilibria with the largest and smallest shares of individuals giving a particular answer. To simplify expressions we will focus on the case where $g(x) = \gamma x^2$.

Assumption 1. *Function $h(\cdot)$ is such that its inverse, $h^{-1}(\cdot)$, is differentiable and satisfies $\frac{dh^{-1}(y)}{dy} \leq \frac{1}{2\beta\sqrt{\frac{y}{\beta}+\sigma^2}}$ for some $\beta > \gamma$.*

This assumption guarantees that the cost of lying is steeper than a certain linear function for x_i close to q and than a certain quadratic function for large x_i. It is satisfied, for example, for $h(x) = (\gamma + \varepsilon)(x^2 + 2\sigma x)$ for $\varepsilon > 0$.

Proposition 2. *There is a unique equilibrium which is characterized by a cutoff z, such that individuals with $x_i > z$ choose $d_i = Y$ while individuals with $x_i < z$ choose $d_i = N$. Moreover, if $q > \mu$, then $z > q$, and if $q < \mu$, then $z < q$.*

Suppose now that $q > \mu$. The cutoff z is decreasing in μ and is increasing in σ and q. The equilibrium share of individuals choosing $d_i = Y$ is increasing in μ and decreasing in q; the effect of an increase in σ is ambiguous.

The first part of Proposition 2 highlights the effect of social stigma: fewer people would admit holding a minority belief than the number of people actually holding it, because some types would cave in to social pressure and misstate their preferences. The equilibrium cutoff z is found as the intersection of two curves (see Figure 2). The first one, $B_i = -C_i(Y) + C_i(N)$, captures the relative benefit of answering Y rather than N; this curve is upward sloping, because types with more right position find it easier to answer Y and costlier to answer N. The second one captures the difference in social costs, $S_i(Y) - S_i(N)$ in case the audience believes the individual has type above z rather than below $\dot{z}$. This cost is increasing in z:

a higher z means that answering Y implies that one has a more extreme position, whereas N becomes a more "normal" one. Assumption 1 guarantees that the two curves intersect exactly once and that the first is steeper than the second, implying the comparative statics results from the second part of Proposition 2.

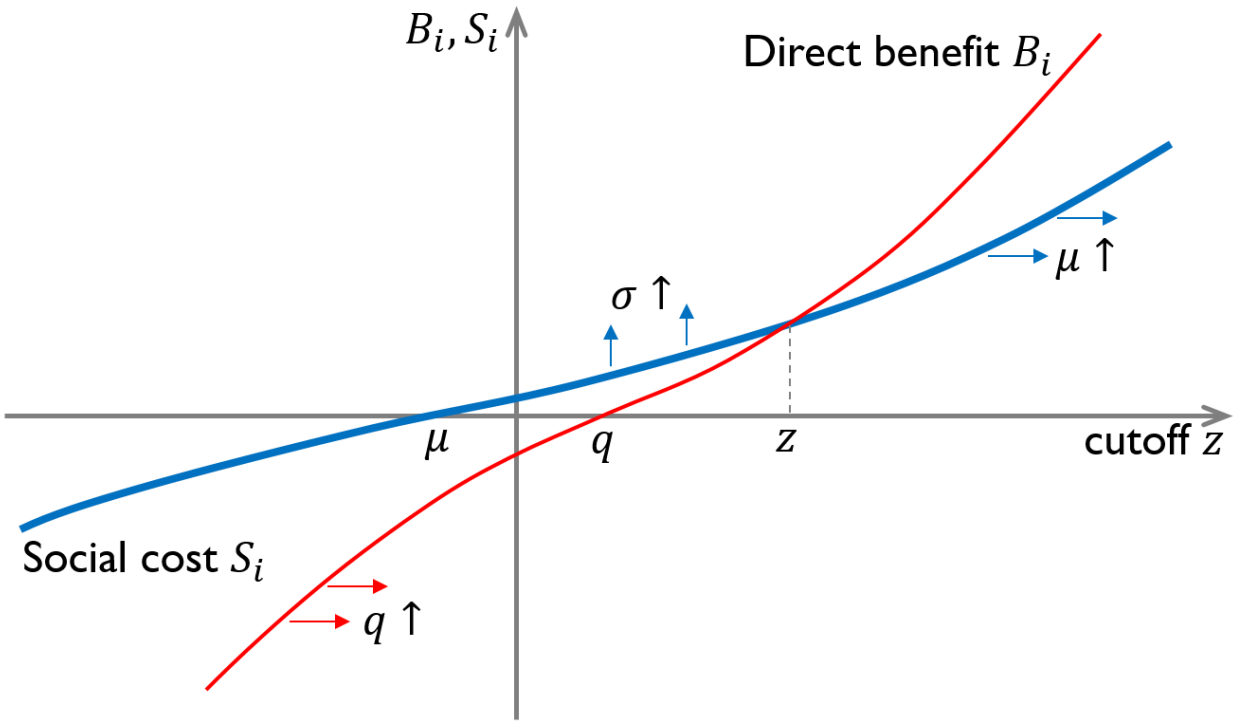

Figure 2: Red/thin line depicts direct benefits (absence of lying costs) of admitting that the individual's type z is above q. Blue/thick line depicts social costs of appearing above z rather than below z for an individual of type z. The arrows illustrate shifts of the curves in response to increases in q, μ, and σ.

As q increases, the share of people answering Y becomes smaller, because agreeing that one's x_i exceeds a high q implies that the individual has extreme views, leading to high social costs. The comparative statics with respect to μ is intuitive as well: an increase in μ implies a shift in the distribution to the right, which means that being xenophobic is more socially acceptable. This increases propensity of individuals of any given type to agree with the statement, so not only does the share of individuals exceeding a given cutoff go

up, but also the cutoff for answering Y goes down, thereby unambiguously increasing the share of individuals choosing Y. The effect of an increase of σ, however, is ambiguous, which might seem counterintuitive given a higher share of xenophobes. However, recall that a mean-preserving spread implies that the share of people despising xenophobes also goes up, and in the end of the day the social pressure is determined by the position of the median individual, which is unchanged. When this individual is moderate and thus dislikes xenophobes, admitting to being a xenophobe is costlier. As a result, even though the share of individuals above any given cutoff above μ is increasing in σ, the equilibrium cutoff itself is increasing in σ as well, thus implying an ambiguous prediction. This ambiguous prediction is consistent with the empirical finding of a noisy effect (see Section 3.1.3, Table 7, and Section 3.2, Table 8). Of course, if the lying cost curve is very steep, the effect of cutoff change is small and a higher σ would increase the share of people answering Y.

5 Conclusion

We study the causal effect of exposure to social media on ethnic hate crimes and xenophobic attitudes in Russia using exogenous variation in initial penetration of social media. We find that higher penetration of social media increases ethnic hate crime. This effect is stronger in cities with a higher baseline level of nationalist sentiment as well as for crimes with multiple perpetrators. The latter finding suggesting that one of the mechanisms behind the effect of social media is through an increase in coordination (as in Enikolopov et al. (forthcoming).

Using a national survey on xenophobic attitudes we further show that social media penetration also had a persuasive effect, especially on young individuals and those with low levels of education. Our design also allows us to investigate whether social media reduced the perceived acceptability of expressing xenophobic attitudes in a survey. We do not find evidence of such decrease – the effect is, if anything, an increase, albeit not significant. We show that

all our results are consistent with a simple model where social media increase individuals' exposure to like-minded individuals, thereby increasing polarization, but inconsistent with a mere shift in opinions towards more xenophobia.

These findings contribute to growing body of evidence that social media is a complex phenomenon that has both positive and negative effects on the welfare of people (see also Allcott et al, 2019), which all have to be taken into account when discussing policy implications of the recent changes in media technologies.

It is important to note that some of our results should be interpreted with caution. First, the problem of weak instruments is an important concern. Even though weak instrument robust methods allow us to get reasonable estimates for our main findings, in most cases power issues prevent us from interpreting the lack of significant results as null effects, due to large confidence intervals, or from studying triple interaction effects to further differentiate between mechanisms. Second, and relatedly, we were only able to conduct our survey experiment in 2018, when the initial shock to social media penetration had already largely dissipated. It is quite possible that we could have learned more about individual and social mechanisms behind the effect if we had conducted our study earlier.

Our paper also hints at promising directions for future research. One interesting question is to find more direct evidence on the effect of social media on polarization and see under which conditions social media may contribute to moderation. More generally, it would be interesting to understand the factors that determine opinion formation. For example, we find evidence consistent with young and low-educated individuals being more impressionable than older or higher-educated ones, but it is an open question which individuals and groups are more likely to be influenced, by whom, and why. Finally, it would be interesting to provide direct evidence on how social media facilitates coordination in practice, by both analyzing the text content in social media forums and understanding how online discussions may lead to offline interactions as well.

References

Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun, "The Power of the Street: Evidence from Egypt's Arab Spring," *Review of Financial Studies*, 2018, *31* (1), 1–42.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya, "Radio and the Rise of the Nazis in Prewar Germany," *Quarterly Journal of Economics*, Nov 2015, *130* (4), 1885–1939.

Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari, "The European Trust Crisis and the Rise of Populism," *Brookings Papers on Economic Activity*, 2017, *Fall*, 309–382.

Allcott, Hunt and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 2017, *31* (2), 211–36.

_ , Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow, "The Welfare Effects of Social Media," Technical Report, National Bureau of Economic Research 2019. Mimeo.

Andrews, Donald W. K., "Identification-Robust Subvector Inference," Working Paper 2017.

Andrews, Isaiah, James Stock, and Liyang Sun, "Weak Instruments in IV Regression: Theory and Practice," *Annual Review of Economics*, 2019 2019, *11*, 727–753.

Blair, Graeme and Kosuke Imai, "Statistical Analysis of List Experiments," *Political Analysis*, 2012, *20* (1), 4777.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler, "A 61-Million-Person

Experiment in Social Influence and Political Mobilization," *Nature*, September 2012, *489* (7415), 295–298.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro, "Greater Internet Use is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups," *Proceedings of the National Academy of Sciences*, 2017, *114* (40), 10612–10617.

Bursztyn, Leonardo and Robert Jensen, "How Does Peer Pressure Affect Educational Investments?," *The Quarterly Journal of Economics*, 2015, *130* (3), 1329.

_ , Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao, "Status Goods: Experimental Evidence from Platinum Credit Cards in Indonesia," *Quarterly Journal of Economics*, 2018, *133* (3), 1561–1595.

_ , Georgy Egorov, and Stefano Fiorin, "From Extreme to Mainstream: The Erosion of Social Norms," working paper 2019.

Cantoni, Davide, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement," *Quarterly Journal of Economics*, may 2019, *134* (2), 1021–1077.

Chaudhuri, Saraswata and Eric Zivot, "A New Method of Projection-based Inference in GMM with Weakly Identified Nuisance Parameters," *Journal of Econometrics*, 2011, *164* (2), 239 – 251.

Coutts, Elisabeth and Ben Jann, "Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)," *Sociological Methods & Research*, 2011, *40* (1), 169–193.

Dasaratha, Krishna, Benjamin Golub, and Nir Hak, "Social Learning in a Dynamic Environment," Technical Report 2019.

DeGroot, M. H., "Reaching a Consensus," *Journal of the American Statistical Association*, 1974, *69*, 118–121.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier, "Testing for Altruism and Social Pressure in Charitable Giving," *Quarterly Journal of Economics*, 2012, *127* (1), 1.

_ , _ , _ , and Gautam Rao, "Voting to Tell Others," *Review of Economic Studies*, 2017, *84* (1), 143.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova, "Social Media and Protest Participation: Evidence from Russia," *Econometrica*, forthcoming.

_ , _ , _ , and Leonid Polishchuk, "Social Image, Networks, and Protest Participation," Technical Report 2017.

_ , Maria Petrova, and Konstantin Sonin, "Social Media and Corruption," *American Economic Journal: Applied Economics*, 2018, *10* (1), 150–74.

Enke, Benjamin, "Moral Values and Voting," working paper 2019.

Glynn, Adam N., "What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment," *Public Opinion Quarterly*, 01 2013, *77* (S1), 159–172.

Golub, Benjamin and Evan Sadler, "Learning in Social Networks," in Yann Bramoullé, Andrea Galeotti, and Brian Rogers, eds., *The Oxford Handbook of the Economics of Networks*, Oxford University Press, 2016, pp. 504–542.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya, "3G Internet and Confidence in Government," Technical Report, mimeo, Paris School of Economics 2019.

Imai, Kosuke, "Multivariate Regression Analysis for the Item Count Technique," *Journal of the American Statistical Association*, 2011, *106* (494), 407–416.

Manacorda, Marco and Andrea Tesei, "Liberation Technology: Mobile Phones and Political Mobilization in Africa," Technical Report 2016.

Mosquera, Roberto, Mofioluwasademi Moffii Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie, "The Economic Effects of Facebook," *Available at SSRN 3312462*, 2018.

Müller, Karsten and Carlo Rasmus Schwarz, "Fanning the Flames of Hate: Social Media and Hate Crime," *Available at SSRN: https://ssrn.com/abstract=3082972 or http://dx.doi.org/10.2139/ssrn.3082972*, 2018.

_ **and** _ , "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment," *Available at SSRN: https://ssrn.com/abstract=3149103*, 2019.

Perez-Truglia, Ricardo and Guillermo Cruces, "Partisan Interactions: Evidence from a Field Experiment in the United States," *Journal of Political Economy*, 2017, *125* (4), 1208–1243.

Qin, Bei, David Strömberg, and Yanhui Wu, "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda," *Journal of Economic Perspectives*, February 2017, *31* (1), 117–40.

_ **,** _ **,** **and** _ , "Social Media, Information Networks, and Protests in China," working paper 2019.

Raghavarao, Damaraju and Walter T. Federer, "Block Total Response as an Alternative to the Randomized Response Method in Surveys," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1979, *41* (1), 40–45.

Settle, Jaime E., *Frenemies: How Social Media Polarizes America*, Cambridge University Press, Aug 2018.

Steinert-Threlkeld, Zachary C., Delia Mocanu, Alessandro Vespignani, and James Fowler, "Online Social Networks and Offline Protest," *EPJ Data Science*, Nov 2015, *4* (1), 19.

Sun, Liyang, "Implementing Valid Two-Step Identification-Robust Confidence Sets for Linear Instrumental-Variables Models," *The Stata Journal*, 2018, *18* (4), 803–825.

Sunstein, Cass R., *Republic.Com*, Princeton, NJ, USA: Princeton University Press, 2001.

_ , *#Republic: Divided Democracy in the Age of Social Media*, Princeton, NJ, USA: Princeton University Press, 2017.

Yanagizawa-Drott, David, "Propaganda and Conflict: Evidence from the Rwandan Genocide," *The Quarterly Journal of Economics*, 11 2014, *129* (4), 1947–1994.

_ **, Maria Petrova, and Ruben Enikolopov**, "Echo Chambers: Does Online Network Structure Affect Political Polarization?," mimeo 2019.

Table 1. Number of Victims by Type.

Victims	Freq.	Percent
Ethnic		
Central Asia	325	23.81%
Caucasians	265	19.41%
Blacks	74	5.42%
Russians	63	4.62%
Arabs	33	2.42%
Jews	10	0.73%
Other "non-slavic"	209	15.31%
Other Asians	108	7.91%
Other Ethnicity	85	6.23%
Total Ethnic	770	56.41%
Non-Ethnic		
Youth groups and left-wing groups	402	29.45%
Religious Groups	106	7.77%
Homeless	42	3.08%
LGBT	32	2.34%
Unknown	13	0.95%
Total Non-Ethnic	595	43.59%
Total	1,365	100%

Table 2. Social Media and Hate Crime. Period: 2007-2015.

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011	-0.130	0.238	-0.229	-0.211	0.348*	-0.417	0.350	0.007	0.502
Weak Instrument Robust Confidence 95% Sets	(-1.169; .660)	(-.262; .787)	(-1.265; .558)	(-1.167; .516)	(-.036; .853)	(-1.460; .316)	(-.423; 1.196)	(-.403; .381)	(-.177; 1.393)
	[0.420]	[0.241]	[0.418]	[0.386]	[0.204]	[0.390]	[0.372]	[0.180]	[0.360]
Nationalist Party Support in 2003	2.407	-0.127	2.208	2.810	-0.801	3.307	-1.178	0.057	-2.176
	[2.492]	[1.250]	[2.477]	[2.301]	[0.956]	[2.341]	[2.069]	[0.945]	[2.042]
Log (SPbSU students, one cohort younger)	-0.084	-0.060	-0.060	-0.155**	-0.058*	-0.122*	0.075	-0.015	0.087
	[0.068]	[0.039]	[0.070]	[0.063]	[0.033]	[0.065]	[0.059]	[0.031]	[0.063]
Log (SPbSU students, one cohort older)	0.101	0.065	0.089	0.099	0.008	0.113*	0.014	0.057*	-0.035
	[0.077]	[0.040]	[0.076]	[0.067]	[0.035]	[0.068]	[0.066]	[0.031]	[0.067]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	13.150	13.150	13.150	13.150	13.150	13.150	13.150	13.150	13.150
Effective F-statistics (Montiel Olea and Pflueger 2013)	13.571	13.571	13.571	13.571	13.571	13.571	13.571	13.571	13.571
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109	23.109	23.109
Endogeneity test p-value	0.575	0.484	0.489	0.461	0.163	0.251	0.390	0.949	0.175

Notes: Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Table 3. Social Media and Hate Crime. Specification with Interaction. Period: 2007-2015.

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	12.002***	6.349***	11.605***	10.578***	5.056**	10.282***	10.365***	1.823	9.125**
Weak Instrument Robust Confidence 95% Sets	(4.537; 23.199)	(1.588; 13.491)	(4.120; 22.833)	(3.701; 20.895)	(1.114; 10.971)	(3.304; 20.749)	(3.004; 21.407)	(-1.623; 6.991)	(1.983; 19.839)
	[4.570]	[2.915]	[4.583]	[4.211]	[2.414]	[4.272]	[4.507]	[2.110]	[4.373]
Log (number of VK users), 2011	0.053	0.362	-0.055	-0.046	0.446**	-0.276	0.529	0.051	0.667*
Weak Instrument Robust Confidence 95% Sets	(-.976; .740)	(-.105; 1.062)	(-1.081; .629)	(-.984; .578)	(.050; 1.041)	(-1.215; .351)	(-.201; 1.624)	(-.410; .359)	(-.036; 1.720)
	[0.420]	[0.286]	[0.419]	[0.383]	[0.243]	[0.383]	[0.447]	[0.188]	[0.430]
Nationalist Party Support in 2003	5.384	1.168	5.534*	4.978*	0.180	5.633*	2.214	0.509	1.137
	[3.298]	[1.527]	[3.260]	[2.930]	[1.281]	[3.006]	[2.557]	[1.096]	[2.504]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts of SPbSU students, older and younger and their interaction with Nationalistic Party Support, 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351
Weak-instrument-robust F-stat for the coefficients of interest	5.759	5.463	5.640	5.491	6.193	6.793	6.246	0.842	6.001
Weak-instrument-robust p-value for the coefficients of interest	0.056	0.065	0.060	0.064	0.045	0.033	0.044	0.656	0.050
Endogeneity test p-value	0.302	0.204	0.291	0.358	0.178	0.218	0.077	0.713	0.066
Full Effect at minimal level of Nationalist Party Support	-0.522	0.057	-0.611	-0.554	0.204	-0.769*	0.032	-0.036	0.229
p-value for the effect at minimum	.255	0.831	.176	.173	.35	.062	0.939	.862	0.573
Full Effect at maximum of Nationalist Party Support	2.584**	1.701**	2.392**	2.184**	1.512**	1.893*	2.715**	0.436	2.591**
p-value for the effect at maximum	0.017	0.027	0.028	0.032	0.021	0.064	0.023	0.380	0.024

Notes: Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Table 4. Social Media and Hate Crime. Specification with Interaction. Period: 2004-2006.

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethnic hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrato*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	-1.898	-0.165	-1.732	-1.321	-0.165	-1.156	-0.576		-0.576
Weak Instrument Robust Confidence 95% Sets	(-5.311; .378)	(-.749; .127)	(-5.065; .489)	(-4.583; .853)	(-.749; .127)	(-4.326; .957)	(-2.094; .436)	.	(-2.094; .436)
	[1.393]	[0.179]	[1.360]	[1.331]	[0.179]	[1.294]	[0.619]		[0.619]
Log (number of VK users), 2011	0.018	0.014	0.005	0.145	0.014	0.132	-0.127		-0.127
Weak Instrument Robust Confidence 95% Sets	(-.258; .433)	(-.021; .066)	[-.271; .280)	(-.075; .476)	(-.021; .066)	(-.086; .459)	(-.376; .039)	.	(-.376; .039)
	[0.169]	[0.021]	[0.168]	[0.135]	[0.021]	[0.133]	[0.101]		[0.101]
Nationalist Party Support in 2003	-0.954	-0.217	-0.737	-1.307*	-0.217	-1.091	0.353		0.353
	[0.772]	[0.213]	[0.717]	[0.769]	[0.213]	[0.695]	[0.370]		[0.370]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Cohorts of SPbSU students, older and younger and their interaction with Nationalistic Party Support, 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes
Observations	625	625	625	625	625	625	625		625
Kleibergen-Paap F-statistics	6.351	6.351	6.351	6.351	6.351	6.351	6.351		6.351
Weak-instrument-robust F-stat for the coefficients of interest	3.041	0.970	2.578	3.848	0.970	3.330	1.577		1.577
Weak-instrument-robust p-value for the coefficients of interest	0.219	0.616	0.275	0.146	0.616	0.189	0.455		0.455
Endogeneity test p-value	0.093	0.468	0.141	0.094	0.468	0.140	0.431		0.431
Full Effect at minimal level of Nationalist Party Support	0.109	0.022	0.088	0.209	0.022	0.187	-0.100		-0.100
p-value for the effect at minimum	.471	.423	.554	.116	.423	.14	.234		.234
Full Effect at maximum of Nationalist Party Support	-0.382	-0.021	-0.361	-0.133	-0.021	-0.112	-0.249		-0.249
p-value for the effect at maximum	0.340	.504	.364	0.698	.504	.743	.245	.	.245

Notes: Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Table 5. Social Media and Ethnic Hostility, Elicited from List Experiment.

	Number of options in List Experiment						
Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dislike Other Ethnicities Option in List Experiment (LE) x Log (Number of VK users, 2011)	0.075**	0.109*	0.043	0.164***	-0.008	0.105***	0.050
Weak Instrument Robust Confidence 95% Sets	(.009; .208)	(-.005; .222)	(-.039; .247)	(.071; .257)	(-.091; .075)	(.026; .224)	(-.055; .207)
	[0.041]	[0.069]	[0.050]	[0.057]	[0.051]	[0.049]	[0.064]
Log (Number of VK users, 2011)	-0.053	-0.001	-0.080	0.017	-0.085	0.066	-0.067
	[0.167]	[0.277]	[0.189]	[0.228]	[0.220]	[0.191]	[0.253]
Dislike Other Ethnicities Option in LE	0.203**	0.110	0.293**	-0.019	0.422***	0.087	0.310**
	[0.101]	[0.173]	[0.123]	[0.131]	[0.130]	[0.119]	[0.157]
Nationalistic Party Support, 2003	-0.832	-1.227	-0.363	-1.390	-0.045	0.120	-1.477
	[1.037]	[1.399]	[1.492]	[1.716]	[1.310]	[1.299]	[1.555]
Dislike Other Ethnicities Option in LE x Vote share of nationalistic party, 2003	1.040	0.680	1.032	0.526	0.762	0.061	2.087
	[1.195]	[2.177]	[1.431]	[1.748]	[1.355]	[1.501]	[1.989]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,447	2,118	2,329	1,954	2,493	2,164	2,283
Kleibergen-Paap F-statistics	4.541	4.366	4.507	4.469	4.445	4.559	4.012

Notes: Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a respondent. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census, and SPbSU older and younger student cohorts.

Table 6. Social Media and Ethnic Hostility, Inferred from List Experiment. City Level.

Subsample:	List Experiment elicited hostility						
	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (Number of VK users, 2011)	0.123***	0.158**	0.050	0.204***	-0.008	0.210***	0.099
Weak Instrument Robust Confidence 95% Sets	(.045, .208)	(.026, .290)	(-.042, .151)	(.099, .334)	(-.136, .107)	(.080, .353)	(-.022, .219)
	[0.041]	[0.070]	[0.049]	[0.062]	[0.062]	[0.069]	[0.064]
Nationalistic Party Support, 2003	1.486	1.058	2.725	1.695	-0.362	1.444	2.912
	[1.522]	[2.700]	[1.896]	[1.953]	[2.178]	[2.500]	[2.188]
Observations	124	116	122	124	111	121	116
Kleibergen-Paap F-statistics	78.994	74.394	81.499	78.994	56.186	73.944	67.506
Effective F-statistics (Montiel Olea and Pflueger 2013)	105.021	98.222	103.035	105.021	75.060	97.275	92.187
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109

Notes: Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside.

Table 7. Social Media and Self-Reported Hostility to Other Ethnicities.

	Self-reported hostility to other ethnicities						
Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. VK and self-reported hate.							
Log (Number of VK users, 2011)	-0.113	-0.200	-0.058	-0.180	-0.087	-0.333	0.159
Weak Instrument Robust Confidence 95% Sets	(-.507; .251)	(-.706; .269)	(-.559; .370)	(-.802; .201)	(-.581; .446)	(-.986;-.016)	(-.292; 1.132)
	[0.147]	[0.189]	[0.188]	[0.203]	[0.199]	[0.188]	[0.240]
Nationalistic Party Support, 2003	-0.010	0.862	-0.501	0.303	-0.178	1.203	-1.548
	[0.805]	[1.134]	[1.115]	[1.302]	[1.127]	[1.129]	[1.251]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,927	927	1,000	853	1,074	943	984
Kleibergen-Paap F-statistics	8.563	8.372	8.486	9.335	8.160	9.561	7.141
Montiel Olea-Pflueger Effective F-stat	8.748	8.533	8.786	9.232	8.095	9.775	6.964
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109
Panel B. VK, self-reported hate, and inferred city-level hate (cities with at least 40 respondents).							
Log (Number of VK users, 2011)	-0.059	-0.257	0.130	0.004	-0.096	-0.195	0.168
Weak Instrument Robust Confidence 95% Sets	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid
	[0.235]	[0.396]	[0.287]	[0.269]	[0.335]	[0.385]	[0.349]
Nationalistic Party Support, 2003	-0.291	0.928	-1.044	-1.359	0.054	0.782	-1.914
	[1.392]	[2.254]	[1.809]	[2.119]	[1.382]	[2.025]	[2.296]
City-level hate to other ethnicities, inferred from LE	0.013	0.016	0.016	0.192**	-0.114	-0.066	0.039
	[0.061]	[0.100]	[0.091]	[0.095]	[0.075]	[0.083]	[0.079]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,382	652	730	569	813	637	745
Kleibergen-Paap F-statistics	2.461	2.023	2.857	4.322	1.927	2.412	2.705
Montiel Olea-Pflueger Effective F-stat	2.500	2.054	2.901	4.386	1.957	2.448	2.747
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109

Notes: Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a respondent. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census, and SPbSU older and younger student cohorts.

Table 8. Social Media and Self-Reported Hostility to Other Ethnicities, FOM.

	Self-reported hostility to other ethnicities						
Subsample:	All	Male	Female	Low Education	High Education	Young	Old
Log (number of VK users), 2011	0.003	-0.065	0.053	0.015	-0.069	-0.091	0.094
Weak Instrument Robust Confidence 95% Sets	(-.248, .254)	(-.442, .199)	(-.150, .373)	(-.242, .313)	(-.513, .197)	(-.448, .173)	(-.122, .453)
	[0.094]	[0.116]	[0.098]	[0.104]	[0.128]	[0.116]	[0.104]
Nationalistic Party Support, 2003	0.487	0.610	0.421	0.455	0.683	0.894	0.134
	[0.522]	[0.615]	[0.559]	[0.568]	[0.596]	[0.653]	[0.537]
Cohorts of SPbSU students, older and younger	-0.002	-0.005	0.010	-0.005	-0.009	0.008	0.001
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,696	12,285	15,411	20,766	6,930	13,571	14,125
Kleibergen-Paap F-statistics	7.897	7.849	7.876	7.798	7.566	8.136	7.368
Montiel Olea-Pflueger Effective F-stat	6.326	6.063	6.500	6.197	6.200	6.179	6.245
Montiel Olea-Pflueger threshold for 10% worst case bias	23.109	23.109	23.109	23.109	23.109	23.109	23.109

Notes: Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence intervals, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a respondent. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census. Indivual-level controls include gender, education categories, dummy for pilot and age categories.

Appendix A

Proof of Lemma 1. First of all, taking the expectation of both sides of (7), and using that $\mathbb{E}x_i^{t-1} = \mathbb{E}x_{-i}^{t-1}$, we get $\mathbb{E}x_i^t = \mathbb{E}x_i^{t-1}$, and therefore $\mathbb{E}x_i^t = \mathbb{E}x_i^0 = \mu$ for each t.

We can iteratively plug in $x_i^{t-1}, x_i^{t-2}, \ldots$ into (7) and use $\mathbb{E}x_{-i}^{t-1} = \mu$ to get

$$\begin{aligned} x_i^t &= (\omega + (1-\omega)\tau)^t x_i^0 \\ &\quad + (1-\omega)(1-\tau)\sum\nolimits_{k=1}^{t} (\omega + (1-\omega)\tau)^{k-1} \mu \\ &\quad + \sum\nolimits_{k=1}^{t} (\omega + (1-\omega)\tau)^{k-1} \varepsilon_i^{t-k+1}. \end{aligned}$$

Since $(\omega + (1-\omega)\tau) \in (0,1)$, the first term converges to 0 in probability as $t \to \infty$. The second term equals

$$(1-\omega)(1-\tau)\frac{1-(\omega+(1-\omega)\tau)^t}{1-(\omega+(1-\omega)\tau)}\mu = \mu - (\omega+(1-\omega)\tau)^t \mu,$$

which converges to μ in probability. Now the last term is a sum of t independent normal variables, and thus the sum is also normal. Its mean is zero, and its variance equals

$$\sum\nolimits_{k=1}^{t} \left((\omega+(1-\omega)\tau)^{k-1}\right)^2 \sigma_\varepsilon^2 = \frac{1-(\omega+(1-\omega)\tau)^{2t}}{1-(\omega+(1-\omega)\tau)^2}\sigma_\varepsilon^2.$$

This latter term converges to σ^2 defined by (8), which implies that the sum converges to $\mathcal{N}(0,\sigma^2)$ in distribution. Since the last term converges to $\mathcal{N}(0,\sigma^2)$ in distribution, and the sum of the first two converges to a constant μ in probability, we have that x_i^t converges to $\mathcal{N}(\mu,\sigma^2)$ in distribution. The comparative statics results are straightforward, which completes the proof. ■

Proof of Proposition 1. We have:

$$\begin{aligned} R_q &= \frac{1}{\sqrt{2\pi}\sigma}\int_q^{+\infty}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx = \left[\frac{x-\mu}{\sigma}=y\right] \\ &= \frac{1}{\sqrt{2\pi}}\int_{\frac{q-\mu}{\sigma}}^{+\infty}\exp\left(-\frac{y^2}{2}\right)dy. \end{aligned}$$

Now,

$$\frac{\partial R_q}{\partial\sigma} = \frac{1}{\sqrt{2\pi}}\frac{q-\mu}{\sigma^2}\exp\left(-\frac{(q-\mu)^2}{2\sigma^2}\right),$$

which is positive if $\mu < q$ and negative otherwise. We furthermore have:

$$\frac{\partial^2 R_q}{\partial\sigma\partial\mu} = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(q-\mu)^2}{2\sigma^2}\right)\left(-\frac{1}{\sigma^2}+\frac{q-\mu}{\sigma^2}\frac{q-\mu}{\sigma^2}\right),$$

which is positive if $|q-\mu| > \sigma$ and negative otherwise. The result follows immediately. ■

Lemma A1. *Let $F(\cdot)$ and $f(\cdot)$ be the c.d.f. and p.d.f. of the standard normal distribution. Then:*

(i) $\phi(x) = \frac{xf(x)}{F(x)(1-F(x))}$ is increasing in x;

(ii) for $x > 0$, $\frac{\phi(x)}{x^2}$ is decreasing in x

(iii) for $x > 0$, $\frac{d\phi(x)}{dx} < \frac{2\phi(x)}{x}$;

(iv) for $x > 0$, $\frac{d\phi(x)}{dx} < 2\sqrt{\phi(x)+1}$.

Proof. (i) The function $\phi(x)$ is odd ($\phi(-x) = -\phi(x)$), so it suffices to prove the statement for $x \geq 0$. Let us prove that $\phi_1(x) = \frac{\sqrt{x}}{F(x)}$ and $\phi_2(x) = \frac{\sqrt{x}f(x)}{1-F(x)}$ are increasing in x for $x > 0$; since both are positive for $x > 0$ this would imply the result.

To prove that $\phi_1(x)$ is increasing in x, consider

$$\frac{d\phi_1(x)}{dx} = \frac{\frac{1}{2\sqrt{x}}F(x) - \sqrt{x}f(x)}{(F(x))^2} = \frac{F(x) - 2xf(x)}{2\sqrt{x}(F(x))^2}.$$

To prove that the numerator is positive, notice that its dervative equals $f(x) - 2f(x) + 2x^2 f(x) = f(x)(2x^2 - 1)$. Thus, the numerator is decreasing on $\left[0, \sqrt{\frac{1}{2}}\right]$ and increasing on $\left[\sqrt{\frac{1}{2}}, \infty\right]$, and since it is positive for $x = \sqrt{\frac{1}{2}}$ (indeed, it equals $F\left(\sqrt{\frac{1}{2}}\right) - \frac{1}{\sqrt[4]{e}\sqrt{\pi}} > \frac{1}{2} - \frac{1}{\sqrt[4]{e}\sqrt{\pi}} > 0$), then it is positive for all $x \geq 0$ and thus $\phi_1(x)$ is increasing in x.

To prove that $\phi_2(x)$ is increasing in x, consider

$$\begin{aligned}\frac{d\phi_2(x)}{dx} &= \frac{\left(\frac{1}{2\sqrt{x}} f(x) - \sqrt{x} x f(x)\right)(1 - F(x)) + \sqrt{x}(f(x))^2}{(1 - F(x))^2} \\ &= \frac{f(x)}{2\sqrt{x}(1 - F(x))^2}\left(\left(1 - 2x^2\right)(1 - F(x)) + 2x f(x)\right).\end{aligned}$$

Denote the last term as $\phi_0(x) = ((1 - 2x^2)(1 - F(x)) + 2x f(x))$; let us prove that it is positive for $x \geq 0$. If $x \leq \sqrt{\frac{1}{2}}$, the first term is positive and the result follows immediately. Suppose that $x > \sqrt{\frac{1}{2}}$, then divide $\phi_0(x)$ by $2x^2 - 1 > 0$ to get $\tilde{\phi}(x) = F(x) - 1 + \frac{2x}{2x^2 - 1} f(x)$; it now suffices to prove that $\tilde{\phi}(x) > 0$ for $x > \sqrt{\frac{1}{2}}$. Notice that $\lim_{x \to +\infty} \tilde{\phi}(x) = 0$ and that $\frac{d\tilde{\phi}(x)}{dx} = f(x) + \frac{\left(2f(x) - 2x^2 f(x)\right)\left(2x^2 - 1\right) - 2x f(x)(4x)}{(2x^2 - 1)^2} = -\frac{6x^2 + 1}{(2x^2 - 1)^2} f(x) < 0$, which means that $\tilde{\phi}(x) > 0$ for $x > \sqrt{\frac{1}{2}}$, and therefore $\phi_0(x) > 0$ for all $x \geq 0$, which establishes that $\phi_2(x)$ is increasing in x. This implies that $\phi(x) = \phi_1(x)\phi_2(x)$ is increasing in x.

(ii) We have $\frac{\phi(x)}{x^2} = \frac{f(x)}{xF(x)(1 - F(x))}$. Let us prove that $\frac{xF(x)(1 - F(x))}{f(x)}$ is increasing in x. Since $F(x)$ is increasing, it suffices to prove that $\phi_3(x) = \frac{x(1 - F(x))}{f(x)}$ is increasing. Differentiating, we get

$$\begin{aligned}\frac{d\phi_3(x)}{dx} &= \frac{(1 - F(x) - x f(x)) f(x) + x f(x) x (1 - F(x))}{(f(x))^2} \\ &= \frac{(1 + x^2)(1 - F(x)) - x f(x)}{f(x)}.\end{aligned}$$

To prove that it is positive, consider $\phi_4(x) = 1 - F(x) - \frac{x f(x)}{1 + x^2}$. We have $\lim_{x \to +\infty} \phi_4(x) = 0$

and

$$\begin{aligned}\frac{d\phi_3(x)}{dx} &= -f(x) - \frac{(1-x^2)(1+x^2)f(x) - 2x^2 f(x)}{(1+x^2)^2} \\ &= -\frac{2f(x)}{(x^2+1)^2} < 0,\end{aligned}$$

which implies that $\phi_4(x) > 0$, and therefore $\phi_3(x)$ is increasing. This, in turn, implies the stated property.

(iii) We have

$$\frac{d\phi(x)}{dx} = \frac{(1-x^2)f(x)F(x)(1-F(x)) + x(f(x))^2(2F(x)-1)}{(F(x)(1-F(x)))^2},$$

which is positive, as we proved in (i). Then

$$\begin{aligned}\frac{d\phi(x)}{dx}\frac{x}{\phi(x)} - 2 &= \frac{(1-x^2)F(x)(1-F(x)) + xf(x)(2F(x)-1)}{F(x)(1-F(x))} - 2 \\ &= \frac{xf(x)(2F(x)-1)}{F(x)(1-F(x))} - x^2 - 1.\end{aligned}$$

We need to prove that the last expression is negative. Notice that $\frac{2F(x)-1}{F(x)} = 2 - \frac{1}{F(x)} \in (0,1)$ for $x > 0$. It thus suffices to prove that $\frac{xf(x)}{1-F(x)} - x^2 - 1 < 0$, which is equivalent to

$$\hat{\phi}(x) = \frac{xf(x)}{x^2+1} - (1-F(x)) < 0.$$

Notice that $\lim_{x\to+\infty}\hat{\phi}(x) = 0$ and that

$$\begin{aligned}\frac{d\hat{\phi}(x)}{dx} &= \left(\frac{1-x^2}{(x^2+1)^2} - \frac{x^2}{x^2+1}\right)f(x) + f(x) \\ &= \frac{2}{(x^2+1)^2}f(x) > 0.\end{aligned}$$

These two facts combined imply that $\hat{\phi}(x) < 0$ for $x > 0$, which implies the required inequality.

(iv) The statement is obviously true for $x < \sqrt{2}$ (it suffices to compute the expressions in a finite number of points). In what follows, we prove that for $x > \sqrt{2}$, $\frac{d\phi(x)}{dx} < 2x + \frac{1}{x}$ and $2x + \frac{1}{x} < 2\sqrt{\phi(x) + 1}$.

Step 1. Let us prove that

$$1 - F(x) < \frac{1}{(1+x^2)^2}. \tag{A1}$$

This inequality obviously holds for small x, and one can easily check that it holds for $x < 4$. Let us show that it also holds for all $x \geq 4$. It suffices to prove that

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{1}{e^{\frac{t^2}{2}}} dt < \int_x^\infty \frac{4t}{(1+t^2)^3} dt.$$

From the fact that

$$e^{\frac{t^2}{2}} > 1 + \frac{t^2}{2} + \frac{t^4}{8} + \frac{t^6}{48}$$

it follows that

$$\frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{1}{e^{\frac{t^2}{2}}} dt < \frac{48}{\sqrt{2\pi}} \int_x^\infty \frac{1}{48 + 24t^2 + 6t^4 + t^6}.$$

It therefore remains to show that for $x \geq 4$,

$$\frac{48}{\sqrt{2\pi}} \int_x^\infty \frac{1}{48 + 24t^2 + 6t^4 + t^6} < \int_x^\infty \frac{4t}{(1+t^2)^3} dt$$

which will be true if

$$\frac{12}{\sqrt{2\pi}} < t + \frac{3t^5 + 21t^3 + 47t}{(1+t^2)^3}.$$

As one can see, the first term increases to infinity as t increases, hence there exists t^* such

that for $t > t^*$ the inequality holds. In particular, this is true for $t \geq 4$, which proves Step 1.

Step 2. Let us show that for $x > 0$, $\phi(x) < 1 + x^2$. This equivalent to

$$F(x)(1 - F(x)) > \frac{f(x)x}{1 + x^2}$$

Define $\psi(x) = F(x)(1 - F(x)) - \frac{f(x)x}{1+x^2}$, clearly $\psi(0) = \frac{1}{4} > 0$. Moreover,

$$\frac{d\psi(x)}{dx} = f(x) - 2F(x)f(x) - \frac{(1 - x^2)f(x)}{1 + x^2} + \frac{2x^2 f(x)}{(1 + x^2)^2} =$$
$$= \left(1 - 2F(x) + \frac{x^4 + 2x^2 - 1}{(1 + x^2)^2}\right) f(x) = 2f(x)\left(1 - F(x) - \frac{1}{(1 + x^2)^2}\right) \underset{\text{(A1)}}{<} 0,$$

which means that $\psi(x)$ is decreasing. Finally, from the fact $\lim_{x\to\infty} \psi(x) = 0$, we can conclude that $\psi(x) > 0, \forall\ x > 0$, which, in turn, implies that $\phi(x) < 1 + x^2$.

Step 3. Let us prove that for $x > \sqrt{2}$

$$\phi(x)F(x) < 1 + x^2 - \frac{1}{1 + x^2}, \tag{A2}$$

which is equivalent to

$$\frac{f(x)}{1 - F(x)} < x + \frac{1}{x} - \frac{1}{x(1 + x^2)}.$$

We have

$$-\frac{d}{dx}\left(f(x)\frac{1 + x^2}{x(2 + x^2)}\right) = f(x)\left(\frac{1 + x^2}{2 + x^2} + \frac{x^4 + x^2 + 2}{x^2(2 + x^2)^2}\right) =$$
$$= f(x)\left(\frac{x^6 + 4x^4 + 3x^2 + 2}{x^6 + 4x^4 + 4x^2}\right) = f(x)\left(1 - \frac{x^2 - 2}{x^6 + 4x^4 + 4x^2}\right).$$

The second term from the expression in the parentheses becomes positive for $x > \sqrt{2}$,

hence the whole expression in the parentheses becomes less than 1. Therefore, for $x > \sqrt{2}$,

$$f(x)\frac{1+x^2}{x(2+x^2)} = \int_x^{+\infty} f(t)\left(1 - \frac{t^2-2}{t^6+4t^4+4t^2}\right)dt < \int_x^{+\infty} f(t)dt = 1 - F(x),$$

which implies

$$\frac{f(x)}{1-F(x)} < \frac{x(2+x^2)}{1+x^2} = x + \frac{1}{x} - \frac{1}{x(1+x^2)}.$$

Step 4. Let us prove that for $x > \sqrt{2}$,

$$\frac{d\phi(x)}{dx} < 2x + \frac{1}{x}, \tag{A3}$$

which is equivalent to

$$\frac{d\phi(x)}{dx} = \frac{\phi(x)}{x}\left((2F(x)-1)\phi(x) + 1 - x^2\right) < 2x + \frac{1}{x}.$$

Using the fact that $\phi(x) < 1 + x^2$, which we proved in Step 2, we have

$$\frac{\phi(x)}{x}\left((2F(x)-1)\phi(x) + 1 - x^2\right) < \frac{1+x^2}{x}((2F(x)-1)\phi(x) + 1 - x^2).$$

It remains to show that

$$(2F(x)-1)\phi(x) + 1 - x^2 < 2 - \frac{1}{1+x^2},$$

and this follows from

$$(2F(x)-1)\phi(x) < F(x)\phi(x) \underset{\text{(A2)}}{<} 1 + x^2 - \frac{1}{1+x^2}.$$

Step 5. Let us prove that for $x > \sqrt{2}$

$$2x + \frac{1}{x} < 2\sqrt{\phi(x) + 1}, \tag{A4}$$

which is equivalent to

$$\phi(x) > x^2 + \frac{1}{4x^2}. \tag{A5}$$

Consider the function $\chi(x) = \phi(x) - x^2$. Clearly, $\chi(0) = 0$, moreover, the this function tends to 1 as $x \to \infty$ is 1. The last property follows from the following consideration. Take the Laurent expansion of $1 - F(x)$ at $x = \infty$:

$$1 - F(x) = f(x)\left(\frac{1}{x} - \frac{1}{x^3} + O\left(\frac{1}{x^5}\right)\right), x \to \infty.$$

From this we have

$$\lim_{x\to\infty} \phi(x) - x^2 = 1.$$

Going back to (A5), we can rewrite it as follows:

$$\chi(x) > \frac{1}{4x^2}.$$

We showed that $\chi(0) = 0$ and the limit of $\chi(x)$ as $x \to \infty$ is 1, whereas the right-hand side is a positive and monotonically decreasing function that tends to 0 as $x \to \infty$. This means that there exists x^* such that for $x \geq x^*$, the inequality (A5) holds, and one can take $x^* = 1 < \sqrt{2}$. This proves Step 5.

Taken together, these steps establish the required inequality. ■

Proof of Proposition 2. First of all, define

$$H(y) = \begin{cases} h(y) & \text{if } y \geq 0; \\ -h(-y) & \text{if } y < 0; \end{cases}$$

then $H(y)$ is a strictly increasing odd function. It is easy to see that the difference in direct costs of an individual with position x_i to give answer Y as compared to N to the question whether x_i exceeds q equals $C_i(Y) - C_i(N) = H(q - x_i)$. Indeed, if $x_i < q$ then saying N is costless whereas the cost of saying Y is $h(q - x_i)$; if $x_i > q$ then saying Y is costless while the cost of saying N is $h(x_i - q) = -H(q - x_i)$, so the difference is $\tilde{h}(q - x_i)$ in this case as well.

Let us show that if in an equilibrium individual i with type x_i weakly prefers $d_i = Y$, then any individual k with type $x_k > x_i$ strictly prefers $d_i = Y$. This follows immediately from that the social cost of the individuals does not depend on their type, and the differences in the direct costs equal $H(q - x_i)$ and $H(q - x_k)$, respectively. Since $H(\cdot)$ is strictly increasing, the difference for agent k is smaller, so the decision $d_i = Y$ involves less cost and the resut follows. This implies, in particular, that every equilibrium must take the form of a cutoff z, with individuals with type $x_i > z$ choosing $d_i = Y$ in equilibrium, wherease those with type $x_i < z$ choosing $d_i = N$.

Let us take a closer look at the social costs $S_i(N)$ and $S_i(Y)$ given the cutoff z. We have

$$\begin{aligned} S_i(N) &= \int_{-\infty}^{\infty} \mathbb{E}_{-i}(g(x - y) \mid x < z) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy \\ &= \int_{-\infty}^{\infty} \frac{\int_{-\infty}^{z} \gamma(x - y)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx}{F\left(\frac{z-\mu}{\sigma}\right)} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy \\ &= \frac{1}{F\left(\frac{z-\mu}{\sigma}\right)} \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) K(x)\, dx, \end{aligned}$$

where the term

$$K(x) = \gamma \int_{-\infty}^{\infty} (x-y)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy$$

captures the social cost an individual whose type is known to be x from interacting with a random individual y. Our assumption that $g(\cdot)$ is quadratic allows us to compute this integral explicitly:

$$\begin{aligned} K(x) &= \gamma \int_{-\infty}^{\infty} \left((x-\mu)^2 + (y-\mu)^2 - 2(x-\mu)(y-\mu)\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy \\ &= \gamma\left((x-\mu)^2 + \sigma^2\right). \end{aligned}$$

Thus,

$$\begin{aligned} S_i(N) &= \frac{\gamma}{F\left(\frac{z-\mu}{\sigma}\right)} \int_{-\infty}^{z} \frac{(x-\mu)^2+\sigma^2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{\gamma}{F\left(\frac{z-\mu}{\sigma}\right)} \left(\int_{-\infty}^{z} \frac{(x-\mu)^2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx + \sigma^2 F\left(\frac{z-\mu}{\sigma}\right)\right) \\ &= \gamma\sigma^2 + \frac{\gamma\sigma^2}{F\left(\frac{z-\mu}{\sigma}\right)} \int_{-\infty}^{\frac{z-\mu}{\sigma}} \frac{t^2}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \\ &= \gamma\sigma^2\left(1 + \frac{F\left(\frac{z-\mu}{\sigma}\right) - \frac{z-\mu}{\sigma} f\left(\frac{z-\mu}{\sigma}\right)}{F\left(\frac{z-\mu}{\sigma}\right)}\right) = \gamma\sigma^2\left(2 - \frac{\frac{z-\mu}{\sigma} f\left(\frac{z-\mu}{\sigma}\right)}{F\left(\frac{z-\mu}{\sigma}\right)}\right), \end{aligned}$$

where we used the fact that $\frac{d}{dx}\left(F(x) - xf(x)\right) = x^2 f(x)$. We can similarly find

$$S_i(Y) = \gamma\sigma^2\left(2 + \frac{\frac{z-\mu}{\sigma} f\left(\frac{z-\mu}{\sigma}\right)}{1 - F\left(\frac{z-\mu}{\sigma}\right)}\right),$$

and therefore

$$S_i(Y) - S_i(N) = \gamma\sigma^2 \frac{\frac{z-\mu}{\sigma} f\left(\frac{z-\mu}{\sigma}\right)}{F\left(\frac{z-\mu}{\sigma}\right)\left(1 - F\left(\frac{z-\mu}{\sigma}\right)\right)} = \gamma\sigma^2 \phi\left(\frac{z-\mu}{\sigma}\right),$$

where $\phi(\cdot)$ is defined in Lemma A1.

In equilibrium, type z is indifferent between choosing Y and N. For this type,

$$\begin{aligned} U_i(Y) - U_i(N) &= -(C_i(Y) - C_i(N)) - (S_i(Y) - S_i(N)) \\ &= H(z-q) - \gamma\sigma^2\phi\left(\frac{z-\mu}{\sigma}\right). \end{aligned}$$

Under Assumption 1, this function $U_i(Y) - U_i(N)$ is increasing in z and has a unique root; this follows from property (iv) of Lemma A1. It is straightforward to check that if $q = \mu$, then this root is $z = q$. Now, since $U_i(Y) - U_i(N)$ is decreasing in q, it must be that for $q > \mu$ we have $z > q$ and for $q < \mu$ we have $z < q$.

Assume now that $q > \mu$. Since $\phi(\cdot)$ is an increasing functions, $U_i(Y) - U_i(N)$ is increasing in μ, and as noted above it is decreasing in q. Furthermore, the latter term may be rewritten as $\gamma(z-\mu)^2 \frac{1}{y^2}\phi(y)$, and by property (ii) of Lemma A1, this term is decreasing in y and therefore increasing in σ, which implies that $U_i(Y) - U_i(N)$ is decreasing in σ. Consequently, the equilibrium cutoff z is increasing in q and σ and decreasing in μ.

These results imply the following about the equilibrium share of types above z, which is equal to $\rho = 1 - F\left(\frac{z-\mu}{\sigma}\right)$. If q increases, then ρ decreases, because z is increasing in q. Similarly, if μ increases, then ρ increases. The comparative statics with respect to σ is ambiguous, because $\frac{z-\mu}{\sigma}$ may increase or decrease (since z is increasing in σ), and one can easily construct examples with with positive and negative effects. ■

Appendix B

Table A1. Summary statistics. City level sample.

Variable	Obs	Mean	Std. Dev.	Min	Max
Log (1+hate crime victims)	625	0.51	0.96	0	4.76
Log (1+hate crime victims, conducted by single perpetrator)	625	0.15	0.42	0	2.56
Log (1+hate crime victims, conducted by multiple perpetrators)	625	0.44	0.92	0	4.74
Log (1+ethnic hate crime victims)	625	0.4	0.84	0	4.36
Log (1+ ethnic hate crime victims, conducted by single perpetrator)	625	0.09	0.32	0	2.56
Log (1+ ethnic hate crime victims, conducted by multiple perpetrators)	625	0.36	0.81	0	4.34
Log (1+non-ethnic hate crime victims)	625	0.23	0.65	0	3.69
Log (1+ non-ethnic hate crime victims, conducted by single perpetrator)	625	0.08	0.29	0	2.2
Log (1+ non-ethnic hate crime victims, conducted by multiple perpetrators)	625	0.18	0.6	0	3.64
Log(1+VK users in a city)	625	9.54	1.33	6.61	13.84
Share of the number of nationalistic voters in 2003 in a population	625	0.05	0.02	0	0.26

Table A2. Social Media and Hate Crime. First stage and Placebo Checks.

	Log(Number of VK users, 2011)	Nationalistic party support in 2003
	(1)	(2)
Log (SPbSU students), same 5-year cohort as VK founder	0.142***	
	[0.039]	
Log (SPbSU students), one cohort younger than VK founder	-0.024	-0.002
	[0.042]	[0.002]
Log (SPbSU students), one cohort older than VK founder	0.051	-0.002
	[0.044]	[0.002]
Nationalistic party support in 2003	4.602***	
	[1.178]	
Log(Number of VK users, 2011)		-0.016
		[0.011]
Socioeconomic city-level controls	Yes	Yes
Observations	625	625

Notes: Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Table A3. Social Media and Hate Crime, Number of Crimes. IV Specification with Interaction. Period: 2007-2015.

	Log (# of hate crimes)			Log (# of ethnic hate crimes)			Log (# of non-ethninc hate crimes)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	11.379***	5.516***	10.971***	10.567***	4.605***	10.188***	6.388**	1.289	5.369**
Weak Instrument Robust Confidence 95% Sets	(5.129; 20.755	(1.445; 11.624	4.890; 20.092	(4.600; 19.516	(1.215; 9.690	(4.315; 18.999	1.395; 13.878	(-1.598; 5.620)	(.839; 12.163)
	[3.827]	[2.493]	[3.723]	[3.653]	[2.076]	[3.596]	[3.057]	[1.767]	[2.773]
Log (number of VK users), 2011	0.081	0.286	0.025	0.005*	0.308	-0.120	0.391	0.101	0.438*
Weak Instrument Robust Confidence 95% Sets	(-.781; .655)	(-.112; .884)	(-.817; .586)	(-.812; .550)	(-.027; .811)	(-.930; .420)	(-.101; 1.129)	(-.155; .358)	(-.002; 1.097)
	[0.352]	[0.244]	[0.344]	[0.334]	[0.205]	[0.331]	[0.301]	[0.157]	[0.269]
Nationalist Party Support in 2003	4.754*	1.182	4.707*	4.761*	0.550	5.038**	0.863	0.177	0.170
	[2.657]	[1.359]	[2.562]	[2.475]	[1.133]	[2.471]	[1.759]	[0.988]	[1.611]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts of SPbSU students, older and younger and their interaction with Nationalistic Party Support, 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	625	625	625	625	625	625	625	625	625
Kleibergen-Paap F-statistics	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351	6.351
Weak-instrument-robust F-stat for the coefficients of inter	7.537	5.425	7.598	7.581	5.685	8.695	5.441	0.976	5.667
Weak-instrument-robust p-value for the coefficients of int	0.023	0.066	0.022	0.023	0.058	0.013	0.066	0.614	0.059
Endogeneity test p-value	0.147	0.213	0.135	0.144	0.215	0.102	0.130	0.697	0.085
Full Effect at minimal level of Nationalist Party Support	-0.465	0.022	-0.501	-0.501	0.087	-0.609*	0.084	0.040	0.181
p-value for the effect at minimum	.201	.926	.152	.133	.645	.064	.776	0.825	.497
Full Effect at maximum of Nationalist Party Support	2.480***	1.449**	2.338**	2.234**	1.279**	2.028**	1.738**	0.373	1.570**
p-value for the effect at maximum	.008	.026	.012	.015	.02	.025	.026	.354	.026

Notes: Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Table A4. Social Media and Hate Crime. IV Specification with Interaction. Different Periods.

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
2007-2009	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	8.226**	2.553*	8.071**	9.202***	3.146**	8.311***	4.538	-0.330	4.743
Weak Instrument Robust Confidence 95% Sets	(1.443; 15.008)	(-.165; 6.631)	(1.331; 14.810)	(2.957; 15.447)	(.644; 6.899)]	(2.273; 14.348)	(-1.605; 13.754)	(-2.424, 1.066)	(-1.382; 13.931)
	[4.152]	[1.664]	[4.126]	[3.824]	[1.532]	[3.696]	[3.761]	[0.855]	[3.750]
Log (number of VK users), 2011	0.133	0.171	0.202	-0.089	0.102	-0.054	0.778***	0.111*	0.740**
Weak Instrument Robust Confidence 95% Sets	(-.551; 1.158)	(-.117; .604)	(-.487; 1.234)	(-.736; .557)	(-.169; .508)	(-.690; .583)	(.201; 1.644)	(-.013; .296)	(.166; 1.600)
	[0.418]	[0.177]	[0.421]	[0.396]	[0.166]	[0.390]	[0.354]	[0.076]	[0.351]
Nationalist Party Support in 2003	0.854	-0.016	0.593	2.540	0.543	2.080	-2.280	-0.685	-0.685
	[1.976]	[0.908]	[1.990]	[1.958]	[0.888]	[1.851]	[2.012]	[0.453]	[1.958]
Weak-instrument-robust F-stat for the coefficients of interest	3.919	2.984	3.509	8.563	5.577	6.812	6.168	3.183	5.800
Weak-instrument-robust p-value for the coefficients of interes	0.141	0.225	0.173	0.014	0.061	0.033	0.046	0.204	0.055
Endogeneity test p-value	0.449	0.371	0.492	0.107	0.157	0.205	0.062	0.223	0.068
Full Effect at minimal level of Nationalist Party Support	-0.262	0.049	-0.185	-0.530	-0.049	-0.452	0.561*	0.126	0.512
p-value for the effect at minimum	.462	.767	.609	.108	.755	.161	.098	.108	.122
Full Effect at maximum of Nationalist Party Support	1.867	0.710	1.904*	1.851*	0.765*	1.699	1.735*	0.041	1.740*
p-value for the effect at maximum	.105	.112	.096	.086	.061	.107	.074	.844	.074
2010-2012									
Log (number of VK users), 2011 x Nationalist Party Support in 2003	10.671***	3.394*	9.319***	7.002**	3.490**	4.249	5.775**	-0.250	6.135**
Weak Instrument Robust Confidence 95% Sets	(4.360; 20.136)	(-.085; 6.873)	[(3.195; 18.504)	[(1.251; 12.753)	(.391; 8.138)	(-1.564; 10.063]	(.563; 13.593)	(-2.603; 2.103)	(1.248; 13.466)
	[3.864]	[2.130]	[3.749]	[3.521]	[1.897]	[3.559]	[3.191]	[1.441]	[2.992]
Log (number of VK users), 2011	0.167	0.199	0.111	0.286	0.263*	0.060	-0.132	-0.031	-0.053
Weak Instrument Robust Confidence 95% Sets	(-.766; .789)	(-.133; .696)	(-.795; .715)	(-.276; 1.128)	(-.028; .700)	(-.483; .602)	(-.940; .407)	(-.333; .170)	(-.808; .451)
	[0.381]	[0.203]	[0.370]	[0.344]	[0.178]	[0.332]	[0.330]	[0.123]	[0.308]
Nationalist Party Support in 2003	5.365*	1.312	4.320	2.592	0.947	1.743	4.014*	0.190	3.685*
	[2.884]	[1.125]	[2.715]	[2.200]	[0.938]	[2.074]	[2.136]	[0.694]	[2.053]
Weak-instrument-robust F-stat for the coefficients of interest	6.908	2.792	5.473	3.599	4.109	1.370	4.390	0.093	4.918
Weak-instrument-robust p-value for the coefficients of interes	0.032	0.248	0.065	0.165	0.128	0.504	0.111	0.955	0.086
Endogeneity test p-value	0.224	0.508	0.353	0.601	0.194	0.910	0.183	0.870	0.102
Full Effect at minimal level of Nationalist Party Support	-0.344	0.036	-0.336	-0.050	0.096	-0.144	-0.409	-0.019	-0.347
p-value for the effect at minimum	.387	.857	.388	.88	.537	.658	.211	.892	.257
Full Effect at maximum of Nationalist Party Support	2.418**	0.914*	2.076**	1.762*	0.999*	0.956	1.086	-0.084	1.241
p-value for the effect at maximum	.011	.093	.023	.053	.051	.292	.185	0.800	.106
2013-2015								\	
Log (number of VK users), 2011 x Nationalist Party Support in 2003	3.476	1.733	2.677	1.954	-0.077	2.720	2.876	1.937	1.200
Weak Instrument Robust Confidence 95% Sets	(-3.944; 8.423)	(-1.073; 5.942)	(-4.320; 7.341)	(-4.861; 6.498)	(-2.826; 1.756)	(-3.824; 7.082)	(-.252; 7.568)	(-.369; 5.397)	(-.975; 4.463)
	[3.029]	[1.718]	[2.856]	[2.782]	[1.122]	[2.671]	[1.915]	[1.412]	[1.332]
Log (number of VK users), 2011	-0.066	0.195	-0.215	-0.075	0.210**	-0.205	0.057	0.018	-0.006
Weak Instrument Robust Confidence 95% Sets	(-.710; .363)	(-.062; .581)	(-.824; .191)	(-.634; .297)	(.064; .427)	(-.752; .160)	(-.406; .366)	(-.302; .232)	(-.343; .219)
	[0.263]	[0.158]	[0.249]	[0.228]	[0.089]	[0.223]	[0.189]	[0.131]	[0.138]
Nationalist Party Support in 2003	1.920	-0.489	2.456	1.380	-1.263	2.500*	0.805	0.663	0.446
	[1.697]	[1.105]	[1.552]	[1.385]	[0.819]	[1.338]	[1.178]	[0.860]	[0.898]
Weak-instrument-robust F-stat for the coefficients of interest	1.477	3.166	2.695	0.902	5.933	4.011	2.574	2.482	0.801
Weak-instrument-robust p-value for the coefficients of interes	0.478	0.205	0.260	0.637	0.051	0.135	0.276	0.289	0.670
Endogeneity test p-value	0.711	0.298	0.406	0.730	0.093	0.315	0.211	0.191	0.643
Full Effect at minimal level of Nationalist Party Support	-0.233	0.112	-0.343	-0.169	0.213*	-0.335*	-0.081	-0.075	-0.064
p-value for the effect at minimum	.396	.556	.158	.438	.066	.095	.704	.635	0.683
Full Effect at maximum of Nationalist Party Support	0.667	0.561	0.349	0.337	0.193	0.369	0.664	0.427	0.247
p-value for the effect at maximum	.365	.129	.627	.632	.397	.595	.132	0.160	.417

Notes: Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Table A5. Social Media and Hate Crime. Specification with Interaction. Period: 2007-2015. Cities with population above 50,000.

	Log (# of victims of hate crime)			Log (# of victims of ethnic hate crime)			Log (# of victims of non-ethninc hate crime)		
	total	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*	*total*	*single perpetrator*	*multiple perpetrators*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (number of VK users), 2011 x Nationalist Party Support in 2003	30.821**	17.622*	26.674**	29.040**	11.554	28.426**	19.431	4.586	12.581
Weak Instrument Robust Confidence 95% Sets	(6.061; 129.863)	(-6.670, 66.205)	(3.112; 109.140)	(6.639; 141.043)	(-18.832; 41.940)	(5.303; 144.040)	(-61.585; 73.442)	(-22.193; 31.366)	(-98.791; 68.268)
	[15.160]	[9.915]	[14.426]	[13.715]	[7.442]	[14.157]	[16.534]	[8.198]	[17.047]
Log (number of VK users), 2011	0.378	0.853	0.325	0.106	0.888	-0.134	1.651	0.193	1.839*
Weak Instrument Robust Confidence 95% Sets	(-2.130; 3.721)	(-.325; 3.798)	(-2.086; 3.540)	(-2.248; 3.245)	(-.051; 3.707)	(-3.200; 2.932)	(-.0505; 6.754)	(-.8220; 1.207)	(.210; 6.724)
	[1.024]	[0.721]	[0.984]	[0.961]	[0.575]	[0.939]	[1.041]	[0.414]	[0.997]
Nationalist Party Support in 2003	-0.648	-3.067	-0.882	0.406	-3.024	0.614	-6.937	-1.014	-7.245
	[5.006]	[3.368]	[4.787]	[4.585]	[2.581]	[4.444]	[5.476]	[1.957]	[5.341]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohorts of SPbSU students, older and younger and their interaction with Nationalistic Party Support, 2003	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	323	323	323	323	323	323	323	323	323
Kleibergen-Paap F-statistics	3.134	3.134	3.134	3.134	3.134	3.134	3.134	3.134	3.134
Weak-instrument-robust F-stat for the coefficients of interest	4.529	5.542	3.317	5.325	5.954	5.181	5.105	0.511	5.497
Weak-instrument-robust p-value for the coefficients of interest	0.104	0.063	0.190	0.070	0.051	0.075	0.078	0.775	0.064
Endogeneity test p-value	0.221	0.118	0.381	0.146	0.134	0.157	0.102	0.816	0.067
Full Effect at minimal level of Nationalist Party Support	-1.097	0.010	-0.951	-1.283	0.335	-1.494	0.721	-0.027	1.237
p-value for the effect at minimum	0.330	0.990	0.390	0.207	0.593	0.145	0.555	0.960	0.313
Full Effect at maximum of Nationalist Party Support	6.877*	4.569*	5.950*	6.230*	3.325*	5.860*	5.748	1.160	4.492
p-value for the effect at maximum	0.053	0.054	0.076	0.057	0.058	0.078	0.127	0.527	0.240

Notes: Robust standard errors in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Table A6. Social Media and Self-reported Ethnic Hostility. IV Specification. City Level.

Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (Number of VK users, 2011)	-0.134	-0.270	-0.146	-0.349	0.110	-0.205	-0.207
Weak Instrument Robust Confidence 95% Sets	(-.002; 2.733)	(-1.845; .954)	(-.019; 2.650)	(.005; 3.935)	(-.883; 1.028)	(.264; 3.250)	(-.884; 2.048)
	[0.181]	[0.227]	[0.208]	[0.225]	[0.212]	[0.212]	[0.208]
Nationalistic Party Support, 2003	-0.339	1.110	0.218	0.627	-1.450	0.892	0.040
	[1.093]	[1.459]	[1.199]	[1.207]	[1.243]	[1.333]	[1.172]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	125	125	114	122	118	122	117
Kleibergen-Paap F-statistics	8.390	7.340	6.840	7.084	7.408	7.220	7.054

Notes: Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census, and SPbSU older and younger student cohorts.

Table A7. Social Media, Nationalistic Party Support, and Ethnic Hostility, inferred from List Experiment

	Number of options in List Experiment						
Subsample:	All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dislike Other Ethnicities Option in List Experiment	-4.18	-7.691	-0.986	-4.436	-2.960	-1.870	-6.318
x Log (Number of VK users, 2011)xNationalistic Party Support in							
Weak Instrument Robust Confidence 95% Sets	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid
	[1.434]	[2.894]	[1.761]	[2.027]	[1.973]	[1.947]	[2.546]
Dislike Other Ethnicities Option in List Experiment	0.083	0.12	0.046	0.172	0.005	0.105	0.062
x Log (Number of VK users, 2011)							
Weak Instrument Robust Confidence 95% Sets	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid
	[0.040]	[0.066]	[0.051]	[0.057]	[0.055]	[0.048]	[0.063]
Log (Number of VK users, 2011)	-1.477	0.539	-4.277	-5.427	1.100	5.968	-11.571
x Nationalistic Party Support, 2003							
Weak Instrument Robust Confidence 95% Sets	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid
	[4.635]	[6.573]	[5.512]	[6.758]	[4.795]	[5.910]	[12.810]
Log (Number of VK users, 2011)	-0.131	-0.078	-0.159	-0.217	-0.060	0.097	-0.310
Weak Instrument Robust Confidence 95% Sets	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid	entire grid
	[0.196]	[0.264]	[0.244]	[0.295]	[0.209]	[0.181]	[0.462]
Dislike Other Ethnicities Option in List Experiment	0.181*	0.080	0.286**	-0.037	0.387***	0.089	0.276*
	[0.099]	[0.163]	[0.127]	[0.133]	[0.142]	[0.118]	[0.156]
Nationalistic Party Support, 2003	-3.180	-8.328	2.587	-1.871	-3.009	-8.582	5.143
	[4.931]	[6.911]	[6.111]	[6.489]	[5.973]	[5.792]	[11.745]
Dislike Other Ethnicities Option in LE	9.103***	15.819**	2.905	8.651**	6.729	3.501	14.500***
x Vote share of nationalistic party, 2003	[3.077]	[6.370]	[4.064]	[3.961]	[4.480]	[4.510]	[5.273]
Cohorts of SPbSU students, older and younger	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,447	2,118	2,329	1,954	2,493	2,164	2,283
Kleibergen-Paap F-statistics	0.668	0.616	0.692	0.695	0.620	0.683	0.614

Notes: Robust standard errors clustered at a city level in brackets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a respondent. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Table A8. Social Media and Ethnic Hostility, Elicited from List Experiment. City Level.

	List Experiment inferred hate						
Subsample:	All	Male	Female	Low Education	HIgh Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (Number of VK users, 2011) x Nationalistic Party Support, 2003	-8.502**	-8.130*	-5.341	-13.538**	-3.545	-10.217**	-7.058
Weak Instrument Robust Confidence 95% Sets	(-24.620; -5.278)	(-22.220; 5.960)	(-25.825; 2.853)	(-36.944;-8.857)	(-20.624; 4.994)	(-27.683;-1.484)	(-33.097; 3.357)
	[3.948]	[5.751]	[5.017]	[5.732]	[5.228]	[5.347]	[6.377]
Log (Number of VK users, 2011)	0.985**	-0.144	0.784**	1.234**	0.044	1.046**	0.238
Weak Instrument Robust Confidence 95% Sets	(.476; 4.039)	(-1.614; 1.326)	(.351; 3.386)	(.622; 4.904)	(-1.034; 1.481)	(.619; 3.612)	(-.784; 3.306)
	[0.623]	[0.600]	[0.531]	[0.749]	[0.440]	[0.524]	[0.626]
Nationalistic Party Support, 2003	15.319**	21.708	12.310	27.781**	10.981	22.501*	17.611
	[7.665]	[13.297]	[11.566]	[13.002]	[12.810]	[12.420]	[14.509]
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	124	122	116	124	111	121	116
Kleibergen-Paap F-statistics	3.345	2.984	3.214	3.218	2.704	3.229	2.689
Full effect at min	1.444	0.294	1.072	1.963	0.235	1.597	0.618
Full effect at max	0.586	-0.525	0.534	0.598	-0.119	0.567	-0.093

Notes: Robust standard errors clustered at a city level in brackets. Stars for endogenous variables are based on weak instrument robust confidence sets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a city. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls include logarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census, and SPbSU older and younger student cohorts. Basic demographic controls include gender, education categories, and age categories.

Table A9. Social Media and Self-Reported Ethnic Hostility.

Subsample:	Self-reported hate to other ethnicities: All	Male	Female	Low Education	High Education	Young	Old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log (number of VK users), 2011	1.548	0.959	1.697	-0.072	1.907	1.692	4.488
x Nationalist Party Support in 2003							
Weak Instrument Robust Confidence 95% Sets	entire grid	entire grid	entire grid	(.058, ...)	entire grid	(.074, ...)	entire grid
	[2.854]	[3.819]	[3.927]	[5.160]	[3.560]	[3.325]	[8.106]
Log (Number of VK users, 2011)	-0.025	-0.137	0.060	-0.244	0.060	-0.248	0.323
	[0.145]	[0.162]	[0.192]	[0.224]	[0.203]	[0.156]	[0.394]
Nationalistic Party Support, 2003	-1.414	-2.088	0.017	-1.842	0.619	-1.703	-3.765
	[3.298]	[4.169]	[5.079]	[5.391]	[4.297]	[4.188]	[7.476]
Cohorts of SPbSU students, older and younger	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Socioeconomic city-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,927	927	1,000	853	1,074	943	984
Kleibergen-Paap F-statistics	1.402	1.497	1.389	1.217	1.381	1.586	0.877
Weak-instrument-robust F-stat for the coefficients of interest	0.438	0.817	0.321	1.642	0.355	3.240	1.229
Weak-instrument-robust p-value for the coefficients of interest	0.803	0.665	0.852	0.440	0.838	0.198	0.541
Full Effect of VK at the minimum of Nationalistic Party Support	-.109	-.189	-.032	-.24	-.043	-.339	.081
Full Effect of VK at the maximum of Nationalistic Party Support	.047	-.092	.138	-.247	.147	-.168	.529

Notes: Robust standard errors clustered at a city level in brackets, *** p<0.01, ** p<0.05, * p<0.1. Unit of observation is a respondent. Logarithm of any variable is calculated with 1 added inside. Socioeconomic city-level controls includelogarithm of population according to 2010 Russian Census, age cohort controls (the number of people aged 25-29, 30-34, 35-39, 40-44, 45-49, 50 and older, in each city according to 2010 Russian Census), share of population with higher education in each of the age cohorts according to 2010 Russian Census, dummy for regional center, log (average wage in 2011), dummy for the existence of a university in a city, log (Odnoklassniki users in 2014), ethnic fractionalization according to 2010 Russian Census.

Appendix C

Questionnaire

Q0. Which city have you been living in for the last 6 months?
List of cities

Q1. How often do you use social networks?
One answer

1	Not at all **[skip to question 3]**
2	Once a month or less
3	Once a week
4	Every day or almost every day
5	Several times a day
6	I'm using social networks nonstop

Q2. Which of the following social networks do you use?
Several answers possible + rotation

1	VKontakte
2	Facebook
3	Odnoklassniki.ru
4	LiveJournal
5	Twitter
98	Other **(please specify)**

Q3. Which websites do you visit most often?
One answer

1	News and analytics websites
2	Social networks
3	Games and entertainment websites
4	Online stores
5	Search engines
98	Other
99	Unsure

Q4. On social networks, do you use your real name or an alias?
One answer

1	Real name
2	An alias, for privacy concerns
3	An alias, but for a reason other than privacy concerns

Q5. How many friends/followers do you have in social networks?
One answer

1	Less than 10
2	10-100
3	100-250
4	250-500
5	500-1000
6	More than 1000

Q6. Do you agree with the statement "I get a lot of important news from social networks"?
One answer

1	Agree
2	Somewhat agree
3	Somewhat disagree
4	Disagree

Q7. Do you agree with the statement "Social networks help me find people with similar interests"?
One answer

1	Agree
2	Somewhat agree
3	Somewhat disagree
4	Disagree

Q8. Do you agree with the statement "In social networks, people are more sincere than in real life"?
One answer

1	Agree
2	Somewhat agree
3	Somewhat disagree
4	Disagree

Q9. To what extent do you trust information in social networks?
One answer

1	Completely trust **[skip to question 10]**
2	Somewhat trust **[skip to question 10]**
3	Somewhat distrust **[skip to question 11]**
4	Completely distrust **[skip to question 11]**

Q10. Why do you trust information in social networks?
Several answers possible

1	People are more sincere in social networks than in real life
2	In social networks one can find a variety of opinions
3	Certain information is only available in social networks
98	Other reason **(please specify)**

Q11. Why do you distrust information in social networks?
Several answers possible

1	Many users deliberately spread incorrect information
2	Many users unwittingly spread incorrect information
3	Many users play the fool and write rubbish
98	Other reason **(please specify)**

Q12. In social networks, how often do you encounter:
[scale: A. Very often, B Often, C Occasionally, D Rarely, E Never]
Rotation of statements, one answer

1	Personal insults
2	Obviously incorrect information
3	Extremist statements
4	Propaganda of violence
5	Religious propaganda
6	Pornography

Q13. Which modern technology do you use to organize gatherings with friends or acquaintances?
Several answers possible + rotation

1	Yes, video calls (e.g., Skype)
2	Yes, messengers embedded in social networks (VKontakte, Facebook, etc)
3	Yes, standalone messengers (WhatsApp, Telegram, ICQ, etc)
4	Yes, blogs or public posts in social networks
5	Yes, SMS (short text messages sent over the phone)
6	Yes, phone calls

THERE ARE TWO RANDOMIZED CELLS.

CELL 1 [QUESTION Q14_1]

Q14_1. Please think, which of the following statements you agree with. Without telling which particular statements you agree with, please specify the number of statements you agree with.

THE ANSWER IS A NUMBER BETWEEN 0 AND 5, ROTATION

1	Each week I usually read at least one newspaper or magazine
2	I want Russia to be a country with high living standard
3	I know the name of the Chairman of the Constitutional Court of the Russian Federation
4	I feel annoyance or dislike toward some ethnicities
5	Retirement benefits in our country are sufficiently high

CELL 2 [QUESTIONS Q14_2, 15, 16, 17]

Q14_2. Please think, which of the following statements you agree with. Without telling which particular statements you agree with, please specify the number of statements you agree with.

THE ANSWER IS A NUMBER BETWEEN 0 AND 4, ROTATION

1	Each week I usually read at least one newspaper or magazine
2	I want Russia to be a country with high living standard
3	I know the name of the Chairman of the Constitutional Court of the Russian Federation
4	Retirement benefits in our country are sufficiently high

Q15. Do you feel annoyance or dislike toward some ethnicities?
One answer

1	Yes
2	No

Q16. In your opinion, which percentage of the survey participants from your city answered "Yes" to the previous question? If your answer is the most accurate, you will get an additional 100 rubles.
Enter a number with a percentage sign – restrict from 0 to 100

Q17. How certain are you in your answer to the previous question?
SLIDER FROM 0 (COMPLETELY UNCERTAIN) TO 10 (COMPLETELY SURE)

QUESTIONS ON GENDER AND AGE ARE ASKED ON THE TECHNICAL PAGE "CIRCLE", SURVEY RESTRICTED TO PEOPLE 18 – 55 YEARS OF AGE

S3. Please specify your education.
One answer

1	Incomplete secondary
2	Secondary
3	Vocational
4	Incomplete higher
5	Higher
6	Doctorate
99	Not sure

S4. Please specify your occupation (your position).
One answer

1	Director, deputy director
2	Division head (of a branch, shift, department)
3	Specialist with a higher education (medical doctor, teacher, sales manager, engineer, etc)
4	Mid-level employee (secretary, salesperson, security, driver, etc)
5	Creative work (photographer, artist, actor, etc)
6	Small business (owner of a business or individual entrepreneur)
7	Technical or service personnel
8	Worker
9	Military
10	Student
98	Other **(please specify)**

S5. How would you describe your family's current financial well-being?
One answer

1	Not enough money even for food
2	Enough money for food, but purchasing clothes is problematic
3	Enough money for food and clothes, but purchasing a TV, a fridge or a washer would be difficult
4	Enough money for major appliances, but we would not be able to buy a new car
5	Enough money for everything except expensive purchases like a country house or an apartment
6	No material difficulties. Can afford to buy a country house or an apartment if necessary