

# User ethnicity and gender as predictors of attitudes to ethnic groups in social media texts

withdrawn

**Abstract.** Out-group bias in the context of race and ethnicity has been widely studied. However, little research has been done to study this phenomenon online. In this paper we explore how ethnicity and gender of Russian social media users affects their attitude toward other ethnic groups. Our results show that ethnicity of social media users plays a significant role in their attitude towards ethnic groups. On the average social media users tend to experience more positive attitude to the ethnic groups they belong to, but there are some exceptions — Russians and Tatars. Gender also influence the attitude insofar as men expressed stronger negative emotions toward foreign peoples.

**Keywords:** Ethnicity, Social Media, Russia, Out-group Bias, Stereotypes.

## 1 Introduction

Both Russian and international researchers spend a lot of efforts studying ethnicity. Being an extremely debatable topic, this phenomenon was influenced by the revolutionary development of mass media, the key one being the Internet, which gave a new impetus to ethnic researches. The transformations that initiated the development of information technologies and the emergence of a worldwide network made it possible for every active user of this network to participate in the production of ethnic discourse and they also provide an opportunity for studying big data. The Internet, and especially social media, is an important public arena for discussions about various ethnic groups and a place where “new ethnicities” are emerging [13].

Thus, data from the Internet, although they do not fully replace ethno-demographic data from the public opinion polls, are nevertheless a valuable source of information about the actual ethnocultural and ethno-political processes, including ethnic stereotypes. The results of a study conducted in the United States has shown that these processes have real-world consequences [7]. Researchers have found a positive correlation between Internet access and the number of crimes motivated by racial hatred, and this correlation is stronger in areas with higher levels of racism.

Given all the above, it seems to us that data from the Internet can be used to analyze ethnic processes. Moreover, as a data source, the Internet has an advantage over traditional offline media because it blurs the boundaries between the consumer and the creator of the content. Now any user can register on social media sites and write a blog post, thereby gaining his own voice in the process of global content production. This feature of online content production makes it possible to measure public opinion on a wide-range of important social issues, which was hardly possible to analyze using traditional media.

Another feature of the Internet that makes data from this source especially useful for analysis of ethnic processes is the high degree of authors' anonymity. Anonymity reduces the power of social control [19], which is especially important in communication on such sensitive topics as ethnicity.

There are a lot of sociopsychological studies on racial ingroup and outgroup bias [6, 8, 12, etc.], but very few of them touches upon this issue in relation to the Internet and social media in particular. This is an important issue, since among all of the mass media that the global network has given life to, social media provide the most favorable environment for the development of discussions, including on ethnic topics [10].

In this study we are trying to investigate whether there is a connection between ethnicity and gender of Russian social media users and their attitudes toward various ethnic groups they live with as it expressed in the posts they write in social media. Russian social media provide a unique opportunity for this kind of research for three reasons: 1) Russia is a multi-national state with over 194 ethnic groups according to the latest census [16], 2) most of the population speaks Russian language regardless of their ethnicity [17], 3) Russia has quite high level of Internet penetration (73.09% in 2016 which is not much less than, for example, in USA for the same period of time — 76.18%) [18].

## **2 Hypothesis**

The key issue of this study is to determine how the author's ethnic identity in social media affects his/her attitude to different ethnic groups.

Hypothesis 1: author's attitude towards ethnic group mentioned in the text will be significantly more positive if the author belongs to the mentioned ethnic group. The causes of this phenomenon are explained by a fundamental psychological inclination of humans to divide people into two classes — in-group and out-group [3].

Hypothesis 2: the size of the effect expressed in Hypothesis 1 will significantly vary across ethnic groups. This hypothesis is based on the assumption that some cultures are more open to out-groups than others, and this is usually related to the extent to which a group feels its identity or interests to be endangered from the outside [9].

Hypothesis 3: males will express significantly more negative attitudes to the ethnic groups they mention, as compared to females.

## **3 Data**

At the first stage we compile a comprehensive list of ethnonyms used to search for the texts related to ethnic discussions. This list included most common forms of post-Soviet ethnonyms. To ensure representativeness we collect all texts related to ethnic discussions from the period of time from January 2014 to December 2015. The text is considered as relevant to ethnic discussions in case it contained at least one word or bigram from the generated list. The texts were gathered using the social media monitoring service IQBuzz which monitor pages from thousands of websites looking for predefined words.

To reveal the presence of ethnic stereotypes, we created a subsample of 15,000 texts for manual coding. Each ethnonym in this subsample was represented by 75 texts (with the exception of those that were found in fewer texts) and each text was evaluated independently by three people. Coders were asked to answer the questions listed in the Table 1.

Then we removed the texts that cause difficulties in understanding, removed ethnic groups with <100 labels and average the labels. Thus, we got 10364 unique text with 22763 labels on the level of ethnonyms.

#### 4 Inter-coder agreement

As mentioned earlier, we showed each text to three persons. To evaluate intercoder reliability we used the Krippendorff's alpha coefficient [11], which is widely used for these purposes and showed good results [1]. We also considered that this coefficient depends on the level in which the respondent's response was measured.

Krippendorff's coefficients for our questions are presented in the Table 1. By convention the Krippendorff's alpha less than 0.67 indicates insufficient agreement of the coders on the given question to draw reliable conclusions [11]. Our results show that for the most questions Krippendorff's alpha does not reach the necessary threshold, so they were to be excluded from the analysis.

**Table 1.** Inter-coder agreement on different questions

Variable's title	Krippendorff's $\alpha$	Measurement level
What is the overall author's attitude to this group/person? (negative/neutral/positive)	0,89	ethnicity
Is the ethnic group or person portrayed as a victim or an aggressor in interethnic relations? (yes/no/unclear/ irrelevant)	0,87	ethnicity
Is the ethnic group or person portrayed as superior or inferior compared to others? (former/latter/unclear/irrelevant)	0,86	ethnicity
How strongly a general negative sentiment is expressed in the text? (no/weak/strong)	0,82	text
Does the text contain one, several or no ethnonyms?	0,73	text
Does the author belong to the ethnic group s/he is writing about? (yes/no/not mentioned)	0,72	ethnicity
How strongly a general positive sentiment is expressed in the text? (no/weak/strong)	0,70	text
Does the author refer to a concrete representative of an ethnic group or to the group as a whole? (former/latter/unclear)	0,57	ethnicity
Does the text contain the topic of religion?	0,55	text
Does the text contain the topic of history?	0,51	text
Does the text contain the topic of politics?	0,49	text
Is the text? (yes/no/other language)	0,45	text

Does the author call for offline violence against the mentioned ethnic group/person? (no/openly/latently)	0,42	ethnicity
Does the text mention interethnic conflict? (yes/no/ unclear)	0,42	text
Does the text contain the topic of culture?	0,39	text
Does the text contain the topic of economics?	0,38	text
Does the text mention positive interethnic interaction? (yes/no/unclear)	0,38	text
Is the ethnic group or person portrayed as dangerous? (yes/no/unclear).	0,33	ethnicity
Does the text contain the topic of migration?	0,28	text
Does the text contain understandable the topic of humor?	0,27	text
Does the text contain the topic of ethnicity?	0,19	text
Does the text contain any other topic?	0,17	text
Does the text contain the topic of daily routine?	0,17	text

Looking at such deplorable results one may ask why it is so. It should be said that we are not the first who encountered this problem. According to a recent study devoted to measuring the reliability of coding texts for the presence of hate speech in them, the convergence of the encoder responses, measured with the help of Krippendorff's alpha, varies from 0.18 to 0.29 [15]. And the consent of the coders did not depend on whether they were shown the definition of the language of hostility or they were guided only by their own criteria. The authors suggested that the problem lies in the fact that the concept of "hate speech" is a complex concept that has no unambiguous definition (similar conclusions were expressed in other works [4, 5]), and therefore it is necessary to approach its definition more carefully and divide complex concepts into simpler ones. These conclusions can also be applied to our work, since the definition of the language of hostility is closely related to our. Therefore, following the advice of the authors, we divided this concept into several simpler ones. Without it we could expect even lower results of convergence.

## 5 Methods and results

In order to measure attitudes towards ethnicities we used a set of mixed regression models which allowed us to control the factors that can influence the dependent variable. Text ID was specified as random variable, and others which are ethnic group mentioned in the text, gender and ethnicity of the author as fixed. We also tested whether it is necessary to use region of author's residence as a random variable, but ICC was  $< 0.01$  so less than 1% of the variance in the variables is due to the regions.

To check the Hypothesis No.1 we built mixed regression Model 1 with dependent variable "What is the overall author's attitude to this group/person?". Kernel density estimation plot showed that the distribution of this variable is close to normal, so we used linear regression model. As for independent variables we specified one fixed variable which is "Does the author belong to the ethnic group s/he is writing about?" and two random variables — text ID (random intercept), ethnicity (random intercept).

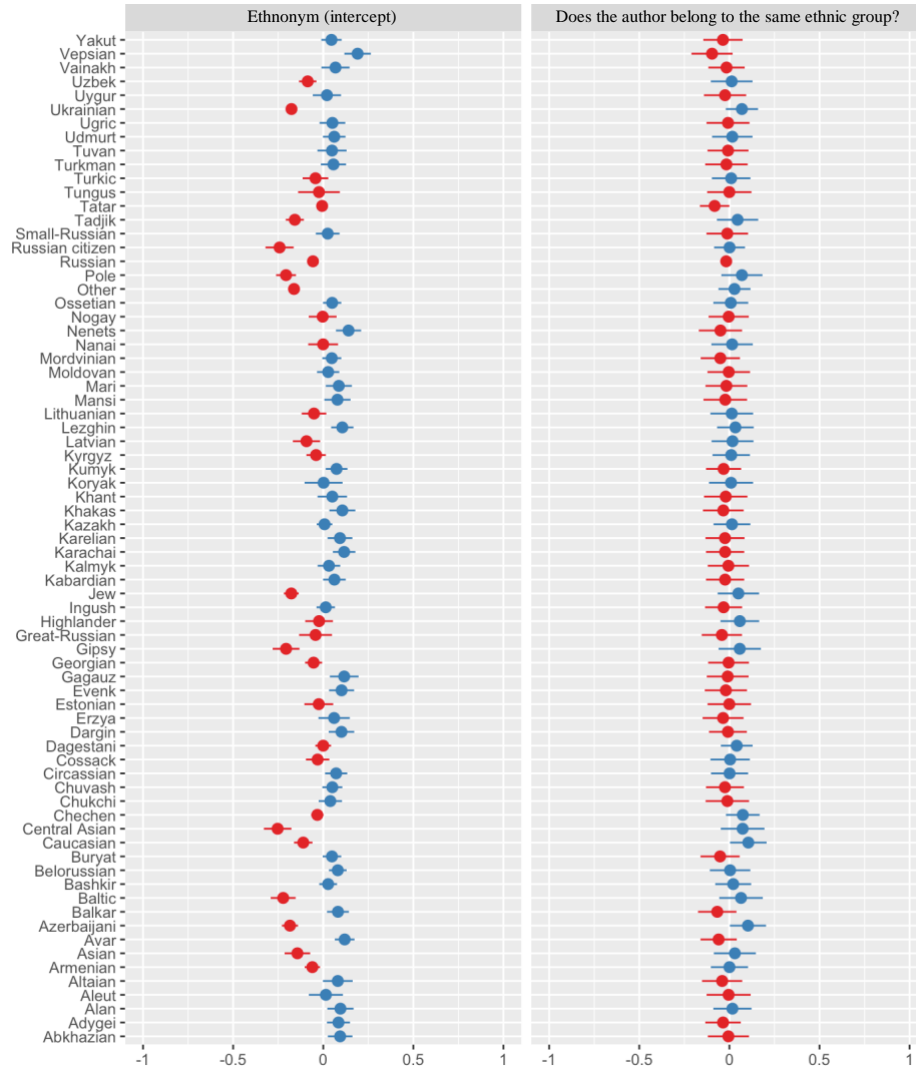
Thus, our model assumes that the average attitude towards different ethnic groups mentioned in different documents may vary.

To test Hypothesis 2, we added to Model 1 the assumption that the relationship between the author's belonging to an ethnic group and his attitude may differ depending on the ethnic group. In other words, we assume the existence of ethnic groups whose members are less positive toward group they represent compared to others. Fig. 1 shows random effects for every ethnonym in the model. The first column shows average attitude towards ethnic group, the second — the strength of author's preference to his/her own ethnic group.

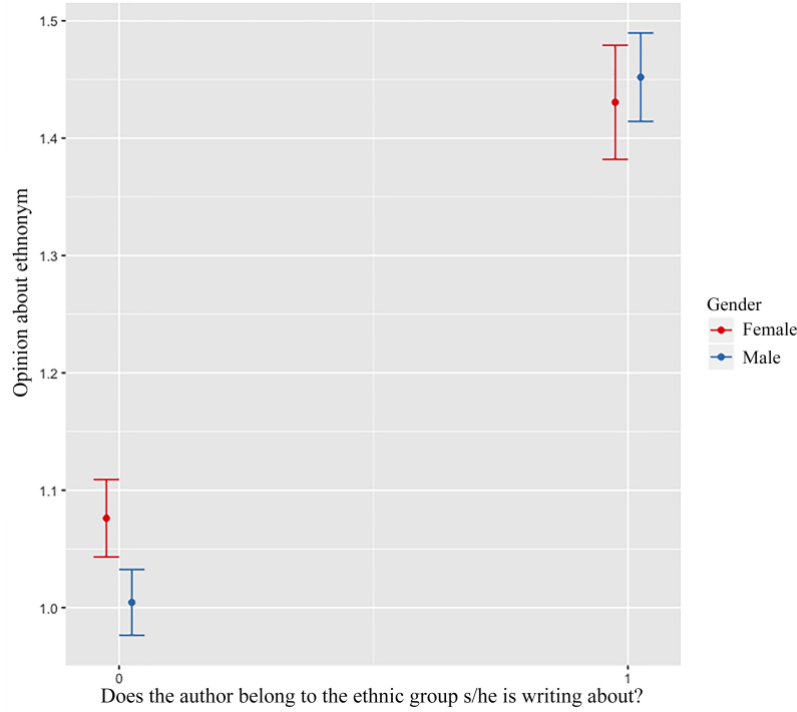
Finally, we created Model 3 with additive and multiplicative interaction between variables "Gender" and "Does the author belong to the ethnic group s/he is writing about?" to test if the author's gender influences his attitude to the ethnic groups. All results are shown in the Table 2.

**Table 2.** Results of regression models with dependent variable  
"What is the overall author's attitude to this group/person?"

Variable	Model 1	Model 2	Model 3
<b>Fixed effects</b>			
Does the author belong to the ethnic group s/he is writing about? (Yes)	0.42 <b>p &lt; 0.001</b>	0.43 <b>p &lt; 0.001</b>	0.35 <b>p &lt; 0.001</b>
Gender (Males)	—	—	-0.07 <b>p &lt; 0.001</b>
Interaction between Gender (Males) and Author's ethnicity (Yes)	—	—	0.09 <b>p &lt; 0.001</b>
<b>Random effects</b>			
Ethnic Group ICC (random intercept)	ICC = 0.05	ICC = 0.05	ICC = 0.05
Text ID (random intercept)	ICC = 0.40	ICC = 0.40	ICC = 0.39
Does the author belong to the ethnic group s/he is writing about? (random slope)	—	—	See Fig. 1
Marginal / conditional R <sup>2</sup>	0.092 / 0.501	0.095 / 0.503	0.098 / 0.497



**Fig. 1.** Model 2: Random effects.



**Fig. 2.** Model 3: interaction between variables “Gender” and “Does the author belong to the ethnic group s/he is writing about?”

## 6 Conclusion

In this paper, we confirmed the hypothesis that the ethnicity of social media users has a significant effect on their attitude towards ethnic groups. This single factor accounts for 9.2% of the variance of the dependent variable in a multilevel regression model, thus Hypothesis 1 is confirmed.

At the same time, if we consider each ethnic group separately, the factor of the author’s ethnicity rarely plays a significant role. Tatars, Russians, Caucasians, and Azerbaijanis are the few groups for which it is significant (Fig. 1). Representatives of Tatars and Russians tend to rate themselves more negatively, and representatives of Caucasians and Azeris more positively. Hereby Hypothesis 2 is partly confirmed.

The last two ethnic groups are part of the peoples of the Caucasus, regarding which in Russia is a widespread stereotype about their great national pride [14], which may explain the observed effect. Interestingly, these two ethnic groups are also characterized by a significantly more negative attitude on average, which is also confirmed by empirical researches [2].

It is more difficult to explain the lower self-esteem of the Tatars and Russians. In the case of Russians, we can refer to a study that revealed negative Russian autostereotypes, such as laziness, laxity and drunkenness [20].

Model 3 also shows that the author's gender plays a significant role in shaping the attitude towards ethnic groups. In average men experience more negative attitude than women. Moreover, if we add the multiplicative effect between the gender of the author and whether he/she is a representative of the ethnic group, it becomes noticeable that for men the fact of this belonging leads to a more polarized attitude compared to women. It can be said that men are more emotionally differentiating their attitude towards ethnic categories depending on their own origin. Hypothesis 3 is confirmed.

## 7 References

1. Antoine, J.-Y. et al.: Weighted Krippendorff's alpha is a more reliable metrics for multicoders ordinal annotations: experimental studies on emotion, opinion and coreference annotation. In: EACL 2014. p. 10 p. , Gotenborg, Sweden (2014).
2. Bodrunova, S.S. et al.: Who's Bad? Attitudes Toward Resettlers From the Post-Soviet South Versus Other Nations in the Russian Blogosphere. *Int. J. Commun.* 11, 0, 23 (2017).
3. Brewer, M.B.: The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *J. Soc. Issues.* 55, 3, 429–444 (1999).
4. Brown, A.: What is hate speech? Part 1: The Myth of Hate. *Law Philos.* 36, 4, 419–468 (2017).
5. Brown, A.: What is Hate Speech? Part 2: Family Resemblances. *Law Philos.* 36, 5, 561–613 (2017).
6. Cassidy, K.D. et al.: The influence of ingroup/outgroup categorization on same- and other-race face processing: The moderating role of inter- versus intra-racial context. *J. Exp. Soc. Psychol.* 47, 4, 811–817 (2011).
7. Chan, J. et al.: The internet and racial hate crime: Offline spillovers from online access. *MIS Q. Manag. Inf. Syst.* 40, 2, 381–403 (2016).
8. Dickter, C.L., Bartholow, B.D.: Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Soc. Cogn. Affect. Neurosci.* 2, 3, 189–198 (2007).
9. Fischer, R., Derham, C.: Is in-group bias culture-dependent? A meta-analysis across 18 societies. *SpringerPlus.* 5, (2016).
10. Kietzmann, J.H. et al.: Social media? Get serious! Understanding the functional building blocks of social media. *Bus. Horiz.* 54, 3, 241–251 (2011).
11. Krippendorff, K.H.: *Content Analysis: An Introduction to Its Methodology.* Sage Publications, Inc, Thousand Oaks, Calif (2003).
12. Levin, S., Sidanius, J.: Social Dominance and Social Identity in the United States and Israel: Ingroup Favoritism or Outgroup Derogation? *Polit. Psychol.* 20, 1, 99–126 (1999).
13. Parker, D., Song, M.: New Ethnicities and the Internet. *Cult. Stud.* 23, 4, 583–604 (2009).
14. Rakacheva, Y.: "Caucasians" in the Russian region: formation of stereotype image and overcoming of prejudice. *Soc. Sociol. Psychol. Pedagogy.* 2, (2012). (in Russian)
15. Ross, B. et al.: Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis. *ArXiv170108118 Cs.* (2017).
16. The Russian Census of 2010: National composition of the population, [http://www.gks.ru/free\\_doc/new\\_site/perepis2010/croc/Documents/Vol4/pub-04-01.pdf](http://www.gks.ru/free_doc/new_site/perepis2010/croc/Documents/Vol4/pub-04-01.pdf).

17. The Russian Census of 2010: Population by ethnithity and ability to speak Russian, [http://www.gks.ru/free\\_doc/new\\_site/perepis2010/croc/Documents/Vol4/pub-04-03.pdf](http://www.gks.ru/free_doc/new_site/perepis2010/croc/Documents/Vol4/pub-04-03.pdf).
18. The Telecommunication Development Sector: Percentage of Individuals using the Internet, [https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2018/Individuals\\_Internet\\_2000-2017.xls](https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2018/Individuals_Internet_2000-2017.xls).
19. Widyanto, L., Griffiths, M.: An Empirical Study of Problematic Internet Use and Self-Esteem. *Int. J. Cyber Behav. Psychol. Learn. IJCBPL*. 1, 1, 13–24 (2011).
20. Kryaklina, T.F.: EHtnicheskie stereotipy kak predmet ehtnopolitologicheskogo ana-liza. *Nauchno-Metodicheskij EHlektronnyj ZHurnal «Koncept»*. 20, 1716–1720 (2014). (in Russian)