INSTITUTO
UNIVERSITÁRIO
DE LISBOA

The Impact of Hate Speech by State Representatives on Attitudes Towards Roma Communities in Bulgaria

Maria Abou Eid-Toteva

Master in Psychology of Intercultural Relations

Dr. Ana Rita Taborda Monteiro Guerra, Integrated Researcher
ISCTE - Instituto Universitário de Lisboa

October, 2024

CIÊNCIAS SOCIAIS
E HUMANAS

Department of Social and Organizational Psychology

The Impact of Hate Speech by State Representatives on Attitudes Towards Roma Communities in Bulgaria

Maria Abou Eid-Toteva

Master in Psychology of Intercultural Relations

Dr. Ana Rita Taborda Monteiro Guerra, Integrated Researcher
ISCTE - Instituto Universitário de Lisboa

October, 2024

Dedicated to my parents.

Acknowledgments

Thank you, Rita, for your unwavering support over the past year! You're not only an incredible supervisor but also an exceptional person. I'll carry the lessons I've learned from you with me for life.

Thank you, Mama and Baba, for instilling in me the values of perseverance, a love for knowledge, and a passion for continuous learning. You've opened every possible door for me in this world!

To Teo, my husband, partner, and best friend—I couldn't have accomplished all of this without you by my side. I love you!

To Ines—I've gained a lifelong friend. Obrigada!

And to myself, for not giving up during the most challenging and emotional two years of my life!

Resumo

O discurso de ódio por parte de representantes do Estado pode moldar atitudes e normas sociais, sendo que as comunidades marginalizadas são geralmente as mais afetadas, como é o caso das comunidades Roma. Neste estudo foram examinados os efeitos da exposição a texto com discurso de ódio em comparação com um discurso neutro nas atitudes, normas sociais, aceitação da discriminação e no apoio ao discurso de ódio contra as comunidades Roma na Bulgária. 174 participantes búlgaros foram distribuídos aleatoriamente por uma de duas condições (discurso de ódio vs. discurso neutro). Os resultados indicaram que não houve um efeito direto da exposição ao discurso de ódio nas atitudes, normas sociais e no apoio ao discurso de ódio; no entanto, análises exploratórias revelaram que a exposição ao discurso de ódio aumentou o desconforto emocional percebido e a perceção da nocividade do discurso de ódio. Análises de mediação indicaram que a exposição a discurso de ódio (vs. não exposição) está indiretamente associada ao apoio ao discurso de ódio e a atitudes negativas face às comunidades Roma através da perceção de desconforto emocional e nocividade do discurso. Ou seja, a exposição ao discurso de ódio aumentou as respostas emocionais (i.e., maior desconforto emocional e nocividade percebida), que por sua vez estão negativamente associadas a atitudes negativas face às comunidades Roma e ao apoio ao discurso de ódio. desconforto emocional e a perceção de nocividade desempenharam um papel importante na explicação do impacto da exposição ao discurso de ódio.

Palavras-chave: Discurso de odio, attitudes, comunidades Roma, normas socais

Abstract

Hate speech by state representatives can shape attitudes and social norms and marginalized communities are usually most affected such as the Roma. This study examined the effects of exposure to hate speech versus neutral speech on attitudes, social norms, and the endorsement of discrimination and the support of hate speech towards the Roma in Bulgaria. 174 Bulgarian participants were randomly assigned to one of the two conditions (hate speech vs. neutral). The results indicated that there was no direct effect of hate speech exposure on attitudes, social norms and support for hate speech; however, exploratory analyses revealed that exposure to hate speech increased perceived emotional discomfort and perceived harmfulness of hate speech. Mediation analysis showed that exposure (vs. not exposure) to hate speech was indirectly related to support for hate speech and negative attitudes toward the Roma via perceived emotional discomfort and perceived harmfulness. That is, exposure to hate speech increased these emotional responses (i.e., higher emotional discomfort and perceived harmfulness), which were then negatively related to negative attitudes towards the Roma and support for hate speech. Emotional discomfort and perceived harmfulness played an important role explaining the impact of hate speech exposure.

Keywords: hate speech, attitudes, Roma communities, social norms

Index

Introduction

Hate speech, which undermines individuals or communities based on factors like race, ethnicity, nationality, or religion, has grave implications for social unity and the continuation of prejudice (Gorenc, 2022). The subtle effects of hate speech are particularly evident when in public discourse, as it reinforces biases and exacerbates the marginalization of vulnerable communities, further deepening social divisions and systemic inequities (Gelber & McNamara, 2015). The impact of hate speech is particularly serious when it targets, dehumanizes marginalized groups and justifies discrimination against them. Regular exposure to hate speech desensitizes people towards aggression, leading them to increased prejudice and isolation within society as highlighted by (Soral et al. 2018). This gradual hate speech desensitization causes individuals' perceptions of acceptable behavior to become less defined as they conform to norms, especially when influential figures endorse similar rhetoric. Social norms play a role in defining what behaviors are considered appropriate within a community, as stated by Brauer and Chaurand (2010). Authority figures play a role in perpetuating the influence of hate speech since the perceived credibility of these individuals can escalate the embrace of language (Tankard & Paluck, 2016).

When influential and political figures publicly express hate speech through their personal views and biases, they shape perceptions of acceptable discourse (Muller & Schwarz, 2020). In this research, we will be focusing on hate speech towards the Roma community in Bulgaria, where such speech boosts harmful stereotypes, fostering an environment that further isolates this vulnerable community, creating a cycle of prejudice and marginalization (Wachs et al., 2022).

The Roma, particularly in East-Central Europe, are one of the most marginalized, socially excluded, and morally denigrated ethnic minority groups (Pogany, 2006). Having their roots in Northern India and settling in Europe more than a thousand years ago, the Roma have endured long periods of discrimination and exclusion (Kende & Krekó, 2020). In Bulgaria, Roma people are a disregarded ethnic minority. Apart from being the most marginalized, they live in rife poverty, limited access to education, and systemic barriers to employment, exacerbating their predicament (Dimitrova et al., 2013). These circumstances stem not only from established systems but also from false stereotypes and xenophobic rhetoric perpetuated by influential figures. The relationship between hate speech and discrimination is bidirectional and cyclic,

where prejudiced verbal expressions by officials enhance the social stigma against Roma, affecting their social identity and community interactions (Cortés, 2021).

This research examines how hate speech by state representatives affects attitudes toward Roma communities and support for hate speech by molding norms that influence perceptions of acceptable behavior and acceptable discourse. Bilewicz and Soral (2020) explain that social norms and public attitudes are shifted when hate speech is promoted by authority figures, thus making biased rhetoric more socially acceptable. A cycle of exclusion and prejudice toward these marginalized and vulnerable communities is reinforced when public views are shifted.

This research aims to understand the consequences of being exposed to hate speech, aiming to provide knowledge to guide efficient campaigns against hate speech and its adverse outcomes.

CHAPTER 1

Literature Review

1.1. Overview of Hate Speech

Discussions on hate speech have been prevalent in many fields. Defining it precisely is a challenge due to its intricate and diverse characteristics. Essentially defined as expressions that belittle or discriminate against individuals or groups based on factors like race, gender, religion, and sexual orientation (Papcunová et al., 2020), hate speech sparks debates and concerns in society. Hate speech contributes to violence and heightened social tension, resulting in a rise in hate crimes following public displays of racist rhetoric, as stated in research findings by Müller & Schwarz (2017. They also highlighted the effects of hate speech in intensifying social divisions and negative attitudes.

The concept of hate speech is closely related to freedom of expression. The classification of hate speech as a category arose from the need to balance protecting communities from harm and freedom of expression (Cohen Almagor, 2011). Over the years, hate speech was mainly linked to discrimination and violent actions. However, nowadays, it encompasses not only those acts but also subtle forms of words and writing that fuel systemic discrimination (Parekh, 2012). For example, Minority groups can be negatively affected by comments or jokes that perpetuate stereotypes and normalize biased attitudes, in daily conversations.

Definitions of hate speech as well as the legal regulations dealing with hate speech differ significantly among regions due to social, cultural and historical backgrounds. Definitions typically focus on language that often provokes animosity, aggression, or bias against groups in public discussions. According to Brown (2017), the lack of a recognized legal definition of hate speech poses challenges in combating it. Institutions and nations have different understandings and law implementations. Expressions of hatred manifest diversely, from acts of violence to subtler forms such as reinforcing damaging stereotypes or depriving certain groups of their basic rights.

Hate speech manifests through comments and prejudiced actions targeting marginalized communities and inciting violence among individuals based on findings by Papcunová et al. (2020). Understanding the significance and implications of hate speech requires examining these classifications, highlighting how such discourse can undermine the dignity of individuals and groups. Identifying and measuring signs of hate speech is a complex and challenging task

to combat. Papcunová et al. (2020) introduced an approach and ways to define hate speech and recognize markers like the use of slurs, personal attacks, and manipulative statements to identify it accurately proving its importance in conversations where such speech aims to marginalize or dehumanize specific groups under the guise of legitimate political discourse.

When used in political discourse, hate speech can target vulnerable communities and portray them as a social threat and a danger to national unity. According to Gelber (2019), hate speech is not merely about voicing dislike or criticism; it involves leveraging one's position of authority to diminish and undermine the rights of a specific group; he highlights that in settings where speakers are in positions of power, their words usually possess added weight and may inflict more harm. This is particularly alarming, as it can normalize discriminatory behavior and attitudes and make them harder to reshape over time.

As noted by Brown (2017), public figures and state representatives who nurture and use hate speech could validate discriminatory attitudes and contribute to a cycle of marginalization and exclusion. Politicians often use rhetoric to mobilize supporters or shift blame onto particular groups for their political gain, amplifying the adverse effects of their speech (Gelber, 2019).

When discussing politics and language use today, it is crucial to distinguish between hate speech and other forms of speech that could be deemed offensive and controversial. Langton (1993) underscores that, when comparing instances of hate speech to other forms of expression, the speaker's credibility can be easily noticed, and the harmful effects of a speech often depend on the reputation of the person delivering it. Gelber (2019) argues that not all offensive language is considered hate speech; for a term to be deemed hate speech, it must meet specific requirements, like perpetuating discrimination and impeding the equal participation of individuals in society. This differentiation plays a role for decision-makers and authorities who must balance safeguarding freedom of expression with preventing harm.

Today, hate speech is a concept that is intricately connected to power dynamics and systemic biases. When it comes to hate speech expressed by political figures, hate speech can be especially harmful as it not only mirrors but prolongs existing societal disparities. Recognizing the different types and signs of hate speech, along with the situations where it emerges, is crucial in devising successful approaches to address it and safeguard the rights and respect of every person.

1.2. Hate Speech Targeting Minority Groups

1.2.1. Consequences and Broader Societal Impact

Victims of hate speech face a stigma that intensifies their isolation and exclusion in society. Individuals who are subjected to hate speech endure heightened stress levels that may result in lasting mental health issues like anxiety disorders and depression (Nadal et al., 2014). Continuous exposure to aggressive language can result in desensitization, lower self-esteem and sense of self, and reduced empathy. Active participation in social interactions and pursuing professional and educational goals become challenging for individuals (Soral et al., 2018). For instance, ethnic minorities often experience increased marginalization that leads to discrimination against them when it comes to work opportunities, service accessibility, and housing according to Anderson (2013).

The harmful effects of hate speech go beyond those directly targeted. They can also impact those who witness or hear such speech. Desensitization—a phenomenon in which frequent exposure results in decreased reactions to negative stimuli—can gradually diminish one's ability to recognize or address the detrimental consequences of hate speech (Soral et al., 2018). Correspondingly, individuals might be less willing to deal with and recognize its outcomes.

Society's increasing tolerance of hate speech is influenced by desensitization. Through repeated exposure to language patterns over time, people may begin viewing them as an aspect of communication rather than something that merits immediate condemnation (Bilewicz & Soral,2020). This altered perception can transform norms, where actions once considered unacceptable become normalized in everyday interactions (Gelber, 2019).

Studies have demonstrated that desensitization can lead individuals towards adopting beliefs and engaging in actions due to exposure over time. One example is the research conducted by Soral et al. (2018), who carried out three studies, including two representative nationwide surveys and one experimental study. In the experimental study, participants were exposed to hate speech in an online environment, where they were asked to evaluate forum posts containing hostile language directed at minorities. The research findings showed that when individuals are repeatedly exposed to hate speech, over time, they tend to become desensitized to its impact and develop biases against the affected communities as a result of this desensitization effect being more accepting of discriminatory behavior and biased thinking (Soral et al., 2018). Furthermore, embracing language could contribute to fostering an environment of bias, where prejudiced attitudes and actions are not just tolerated but

endorsed. As noted by Bilewicz and Soral (2020) and Matsuda (2018), this societal change could lead to divisions among individuals based on race, ethnicity, and religion, potentially undermining societal cohesion and escalating incidents of hate crimes and intergroup conflicts. The enduring repercussions of this divide and dehumanizing rhetoric may manifest as a weakened sense of solidarity, a surge in hate-driven crimes, and an increased likelihood of conflicts between groups, as indicated by (Fasoli et al., 2016).

In essence, directing hate speech towards minority communities has effects that reach beyond the individuals targeted and negatively impact society at large. The emotional and communal damage suffered by victims is worsened by the normalization and desensitization of hate speech among the population. This combined effect does not only sustain prejudice and exclusion, it also jeopardizes the fundamental unity of our society. Dealing with hateful language is crucial not only to safeguard minority communities but also to uphold the principles of acceptance and unity that form the foundation of a more inclusive society.

1.3. The Role of Public and Political Discourse, Social Norms, and Attitudes

1.3.1. Influence of Public and Political Discourse on Hate Speech

Expanding upon the conversation about the damage caused to minority communities by language reveals the importance of examining how public discussions and political discourse can exacerbate these repercussions by influencing societal beliefs and behaviors (Müller & Schwarz, 2020). The direct impact of hate speech towards individuals is grave; however, its implications for societal structure is equally concerning. According to White and Crandall (2017), the presence of hate speech in discussions leads to a shift in norms that support and legitimise discrimination.

Words can lead to tangible effects on society; a study by Müller and Schwarz (2020) found a correlation between hate speech by political figures and an uptick in hate crimes. Places that were subjected to more frequent racist rhetoric from public figures saw a notable surge in hate crimes. By using hate speech, public figures contribute to an atmosphere where discrimination becomes accepted, thereby heightening the chances of violence and exclusion towards particular groups.

Moreover, Crandall et al. (2002) suggest that societal norms shape how prejudices are displayed. When authority figures fail to denounce hate speech, it indirectly allows the public

to express discriminatory views that result in normalizing hate speech. This occurs because individuals often emulate the behaviors of leaders. When hate speech becomes acceptable at the top levels of society, it is more likely to spread among the public. Tankard and Paluck (2016) also highlight in their research that individuals are more likely to adopt discriminatory language and behaviors when authority figures model such rhetoric. According to them, leaders influence shaping social norms.

Bleich (2011) suggests that the media amplifies the influence of political discourse. When politicians use derogatory language in their speech, the media tends to amplify these statements, which can then shape the public's perceptions of minority groups in society.

The study by Muller and Schwarz (2020) further highlights the influence of political figures sensationalizing or repeatedly discussing hate speech issues in discourse; this can shape the way people view and engage with marginalized communities through media coverage.

1.4. The Role of Social Norms and Attitudes

When a group places importance on equality and values diversity in their interactions with groups or individuals, it tends to reduce the acceptance of discriminatory acts in society. However, the opposite outcome will likely happen if biased attitudes and actions become ingrained. According to Paluck (2009), social norms influence behavior in various social settings and communities at large.

The justification suppression model by White and Crandall (2017) delves into how social norms inhibit prejudice by discouraging the expression of discriminatory opinions because of potential social repercussions. These suppressive effects usually fade when hate speech is normalized by influential figures, resulting in an environment where prejudiced opinions are deemed acceptable and expressed without concern for backlash. This means that hate speech does not just mirror prejudices but also plays a role in reshaping social norms by making discrimination more permissible in public conversations.

Álvarez Benjumea and Winter (2020) pointed out that norm shifts can influence people's views toward hate speech, suggesting that when hate speech is accepted as the norm, acts of discrimination are likely to be endorsed and upheld. This cycle perpetuates bias and undermines standards. Tankard and Paluck (2016) emphasise that authority figures are instrumental in shaping norms, according to them authority figures's acceptance or tolerance to this type of speech plays a crucial role in shaping intergroup relations and expressing prejudice. As mentioned previously, Bilewicz and Soral (2020) suggest that when leaders

actively participate in hate speech or fail to denounce it, the public is more inclined to consider such language acceptable and normalize it, making it difficult to perceive its harmfulness when encountering it and desensitizing people.

1.4.1. Impact on Social Cohesion

Hate speech harms people personally and fractures the bonds that unite a society, as Papcunova et al.(2020) highlighted. Such division is driven by hate speech that perpetuates stereotypes and makes prejudice seem acceptable, leading to the exclusion of marginalized communities, more intergroup conflicts, and a rise in social inequalities.

Muller and Schwarz (2020) underline in their research the outcomes of discrimination and social unity when hate speech becomes prominent in political speeches and media narratives. Trust between social groups starts to break down, and society becomes more vulnerable to prejudice, hostility, and violence, as noted by Bilewicz and Soral(2020). This ultimately weakens the core values that promote harmony and inclusivity in our communities, namely embracing differences and fostering unity.

1.5. The Roma in Bulgaria

Minority communities often bear the brunt of hate speech more than others do, as it reinforces stereotypes and fuels discrimination against them. One notable minority group affected by this is the Roma community in Europe, which has been impacted over the years. According to Van Baar (2011), the Roma population originally migrated from India around the 10th century and spread across Europe over time. This migration led to diverse Roma communities adapting to the national context while holding on to their distinct cultural identities.

Throughout history, the Roma community has often been marginalized due to political and societal attitudes toward them (Giroud et al., 2021). The Roma in Bulgaria experience the same kind of marginalization that many Roma communities across Europe face. Negative attitudes toward the Roma are deeply rooted in culture. Often, they come across through public debates and political talks that portray them unfairly and strip them of their humanity.

This situation has led to a rise in discrimination, exclusion from society, and violence against Roma individuals. These unfortunate circumstances have entrenched poverty and marginalization within their community (Dimitrova, 2013).

The Roma communities in Bulgaria encounter challenges beyond economic difficulties—they also confront considerable prejudice driven by political rhetoric and media portrayals that reinforce harmful stereotypes about them as a burden on society linked to crime and poverty rather than as fellow individuals deserving of dignity and fair treatment (Van Baar, 2011).

The normalization of anti-Roma rhetoric in Bulgaria is closely tied to the rise of right-wing populism in Eastern Europe. Nationalist ideologies leverage past biases to establish political influence (Kende et al.,2017). In this context, Romani communities often bear the blame for their challenges, which are seen as jeopardizing national identity and societal peace. Such circumstances have led to a situation where negative attitudes towards the Roma community are not just tolerated socially but are also viewed as politically beneficial (Kende et al., 2017).

The outcomes of language use are serious, as research suggests that they may lead to a rise in violence against Roma people and communities, along with increased exclusion and financial hardship also emphasized by (Cortés, 2021). The acceptance of hate speech directed at Roma individuals in Bulgaria highlights the necessity to tackle the impact of political communication on societal standards and views, as well as the broader effects it has on social unity and human rights.

1.6. The Present Study

This research examines impact of hate speech against the Roma community voiced by state representatives in Bulgaria. Specifically, based on previous research it examines if expositing participants to a message by a state representative expressing hate speech against the Roma in Bulgaria (vs. a neutral condition) increases support of hate speech towards the Roma, negative attitudes toward the community, as well as support for social norms that normalize the expression discrimination towards the Roma (H1). Additionally, we also explored if the detrimental effect of hate speech exposure on attitudes and support for hate speech is mediated by social norms. (H2)

CHAPTER 2

Methods

2.1. Design

This study employed a between-subjects experimental design, with participants randomly assigned to one of two conditions: exposure to hate speech by state representatives or exposure to a neutral text.

2.2. Participants

A power analysis conducted with MedPower (Kenny, 2017) indicated that a minimum of 156 participants was required to detect a small to medium effect size with 80% power. Participants were required to be Bulgarian nationals aged 18 or older. 214 individuals were recruited through convenience sampling via personal networks and university channels at Sofia University, explicitly targeting students from the psychology and sociology departments. After cleaning the dataset 40 participants were excluded due to incomplete responses (e.g., some participants started the survey but did not complete it) or failure to meet the eligibility criteria. This resulted in a final sample of 174 participants who fully completed the survey and met all inclusion criteria. This final sample was used for subsequent analyses.

The participants ranged in age from 18 to 86, with a mean age of 36.83 (*SD* = 13.62)— most participants identified as women, politically centered (*M*=3.60, *SD*=1.34) and regarding their national identification they identified as neutral (*M*= 4.05, *SD*=1.46).

Table 2.1.

Sociodemographic Characteristics of Participants

	Total		Control		Exposure to Hate Speech	
	N = 174		*n* = 85		*n* = 89	
Gender						
Female	119	68.4 %	60	70.6 %	59	66.3 %
Male	53	30.5 %	24	28.2 %	29	32.6 %
I prefer not to answer this question.	2	1.1 %	1	1.2 %	1	1.1 %
Education						
Elementary school	1	0.6 %	1	1.2 %	0	0%
High school	24	13.8%	11	12.9%	13	14.6 %
Bachelor's degree	53	30.5 %	28	32.9 %	25	28.1%

Master's degree/	89	51.1 %	42	49.4%	47	52.8 %
PhD	6	3.4%	3	3.5%	3	3.4%
Chose not to disclose	1	0.6 %	0	0%	1	1.1 %
Subjective income						
Comfortable	85	48.9%	43	50.6 %	42	47.2 %
Coping	76	43.7%	38	44.7%	38	42.7 %
Difficult	9	5.2 %	3	3.5 %	6	6.7 %
Very Difficult	2	1.1 %	0	0 %	2	2.2 %
Don't know	2	1.1%	1	1.2%	1	1.1%

2.3. Procedure

All materials used in this study were approved by the ethics committee of ISCTE (05/2024). As the materials were presented in Bulgarian, they were piloted following translation from English to ensure accuracy. Data collection was conducted online using Qualtrics. Participants were recruited through convenience sampling, social media, and networks involving university students and friends. Participants first encountered the informed consent form, which detailed the researchers' contact information, study duration, and participation's voluntary, anonymous, and confidential nature. Upon agreement, participants proceeded to answer demographic questions. They were then randomly assigned to one of two experimental conditions: exposure to a text containing hate speech by a Bulgarian state representative or a neutral text not containing hate speech. Following exposure to these conditions, participants responded to measures of interest, including hate speech, attitudes towards Roma, social norms, a manipulation check and for exploratory purposes emotional discomfort, perceived harmfulness, and perceived impact of hate speech[1]. After the study, participants were debriefed and thanked for their participation. Details of all materials used are available in the Appendix.

2.4. Materials and Measures

2.4.1. General Demographics

Participants provided demographic information, including their age, gender, highest level of education, subjective income level, political ideology, and national identity (identification with Bulgarians).

2.4.2. Manipulations

Participants were randomly assigned to one of two experimental conditions without prior introduction to the study context. In the **Hate Speech Condition**, participants read an adapted text featuring intensified negative statements about Roma communities attributed to a high-ranking Bulgarian official, which was sourced from actual public statements reported in the media and acknowledged in reports by the European Commission against Racism and Intolerance (ECRI, 2019). Conversely, the **Neutral Condition** presented participants with a text excerpted from a Ministry of Culture report detailing the accomplishments of Georgi Gospodinov, a renowned Bulgarian novelist who won the Booker Prize. This text highlighted cultural achievements and lacked discriminatory content, engaging participants neutrally.

2.4.3. Support for Hate Speech

To assess support for hate speech, participants responded to a 5-item measure developed by Bahador (2020). Items were rated on a 7-point Likert scale, ranging from 1 (strongly agree) to 7 (strongly disagree), encapsulating dimensions such as disagreement with hate speech ("I do not agree with Roma people's viewpoints and values"), stereotyping/generalizing ("Roma people tend to engage in dishonest practices"), negative characterizations/insults ("Roma people are lazy and untrustworthy"), dehumanization/demonization ("Roma people are like a disease to our society"), and incitement to violence ("Roma people should be physically stopped"). (α =.62). Considering the low reliability score we eliminated one item (i.e., I don't agree with Roma people viewpoints and values) and the reliability increased (α =.80). We created a composite score where higher values indicate higher levels of support for hate speech.

2.4.4. Attitudes

Participants completed the Attitudes Toward Roma Scale (ATRS) developed by Kende et al. (2017) to measure attitudes toward Roma communities. This scale comprises various items to assess explicit attitudes and stereotypes. Participants rated their agreement with statements on a 7-point Likert scale from 1 (strongly agree) to 7 (strongly disagree). Example items include: "Roma people tend to commit more criminal acts than other people," "Roma people do not have a positive relationship to work; they are lazy," and "The growing Roma population threatens the security of society." (α =.87). We created a composite score where higher values indicate more negative attitudes.

2.4.5. Social Norms

To evaluate participants' perceptions of social norms related to attitudes toward Roma, we employed a scale adapted from Visintin et al. (2019). This scale measures both descriptive and prescriptive norms concerning societal views of Roma communities. Participants were asked to rate their agreement with two items on a 7-point Likert scale ranging from 1 (strongly agree) to 7 (strongly disagree). The items assess perceptions such as "Bulgarians have negative feelings toward Roma" and "In Bulgaria, it is acceptable to express negative feelings toward the Roma."

Bivariate correlations between the items were calculated which showed that they were moderately (r_s =.30) and significantly ($p < .001$) correlated. We created a composite score where higher values indicate higher support for discriminatory norms.

2.4.6. Emotional Discomfort

Participants' emotional response to the text they read was assessed using one item from the Emotional Discomfort Scale created by Symvoulakis et al. (2022). The item assessed the level of unease and discomfort experienced while reading the manipulation text on a 7-point Likert scale ranging from 1 (not at all) to 7 (extremely uncomfortable).

2.4.7. Perceived Harmfulness

To assess how individuals perceive the impact of hate speech, we used one item derived from Downs and Cowan's (2012) HHSS scale. We assess the degree to which individuals perceive the content of the manipulation texts as harmful using a 7-point Likert scale from 1 (not harmful at all) to 7 (harmful). Participants were asked to rate the harm the text they read can cause in promoting discrimination or violence against the Roma communities.

2.4.8. Manipulation Check

To test the effectiveness of the experimental manipulations, a manipulation check was included immediately following the exposure to the text conditions. Participants were asked to assess the extent to which the text they read contained hate speech on a 7-point Likert scale ranging from 1 (not at all) to 7 (a lot)."To what extent do you consider the text you have read to contain hate speech?"

CHAPTER 3

Results

Statistical analyses were performed using IBM SPSS Statistics (29). To confirm the efficacy of our experimental manipulation, we first conducted an independent samples t-test comparing the perceived hate speech between the two conditions. We then utilized independent samples t-tests to test our primary hypothesis regarding the impact of hate speech exposure on the main dependent variables. Finally, to test the hypothesized mediation effects we used mediation analysis using the SPSS Process macro (Hayes, 2018).

3.1. Preliminary Analysis – Manipulation Check and Test of Assumptions

To evaluate the efficacy of the manipulation, we conducted an independent samples t-test comparing the levels of perceived hate speech between the hate speech condition and the neutral condition. In concordance with what was hypothesized, participants exposed to the hate speech condition (M = 5.65, SD = 1.60) considered the text of the hate speech condition to have significantly more hate speech compared to those in the neutral condition (M = 2.05, SD = 1.63), t (172) = -14.73, p <.001.

Preliminary analyses were also performed to check whether the subsequent statistical test assumptions were met. The Kolmogorov-Smirnov test did not show a normal distribution of perceived hate speech in both the exposed condition, D (85) = .22, p < .001, and neutral condition D (85) = .32, p < .001. This assumption was not met according to the Kolmogorov-Smirnov test. Given the sample size (n > 30), we performed the Central Limit Theorem, which did not allow us to assume an approximately normal distribution of perceptions of hate speech in the two groups. However, the t-test is robust for violations of normality.

The assumption of equal variances was confirmed via Levene's test for equality of variances, F (172) = .03, p = .88. Regarding other study variables, normality assumptions were upheld according to the Kolmogorov-Smirnov tests for attitudes towards Roma in the exposed condition D (89) =.10, p = .03, but not for the neutral condition D (85) =.06, p = .20. Regarding social norms, normality assumptions were upheld according to the Kolmogorov-Smirnov tests for both the exposed condition, D (89) = .12, p = .003, and neutral condition D (85) = .15, p <. 001. However, the histogram indicated that the data was approximately normally distributed. Levene's variance homogeneity test was significant for

social norms, F (172) = .1, p=.76. The Kolmogorov-Smirnov test showed a normal distribution of attitudes on the neutral condition, D (85) =.06, p = .20, but not on the exposed condition, D (89) =.10, p = .03. However, once again, the t-test is robust to normality violations. Levene's variance homogeneity test was significant for attitudes, F (172) = 1.10, p =.30. Regarding support for hate speech, normality assumptions were upheld according to the Kolmogorov-Smirnov tests not for the exposed condition, D (89) = .09, p = .05. Still, for the neutral condition D (85) = .11, p = .02. However, the histogram indicated that the data was approximately normally distributed. Levene's variance homogeneity test was significant for hate speech, F (172) = .001, p=.98.

3.2. Main Analysis – The Impact of Exposure to Hate Speech on Hate Speech Support, Attitudes Towards the Roma, and Social Norms

To assess the impact of the exposure on attitudes towards the Roma, we conducted an independent samples t-test. The independent variable was the condition (exposure to hate speech vs. neutral text), and the dependent variable was the Attitudes Toward Roma Scale (ATRS) score. The results indicated that there was not a significant effect of the condition on attitudes toward Roma t (172) = .81, p =.42. Participants in the hate speech condition *(*M = 4.29*,* SD = 1.57*)* did not show significant differences from the participants in the neutral condition *(*M = 4.47*,* SD = 1.44*)*.

Similarly, a t-test analysis was performed to compare support hate speech between participants exposed to hate speech and those exposed to a neutral text. The results showed no significant differences in the support for hate speech between the two groups, t(172) = -.31, p = .38. Participants in the hate speech condition (M = 3.36, SD = 1.31) did not differ significantly from those in the neutral condition (M = 3.29, SD = 1.32).

A t-test was conducted to compare the means of support for social norms towards the Roma between participants exposed to hate speech and those exposed to neutral text. The results indicated that there were no significant differences in support for discriminatory social norms between the two conditions, t(172) = .81, p = .21. Participants in the hate speech condition (M = 5.81, SD = .15) did not differ significantly from those in the neutral condition (M = 5.99, SD = .17).

Considering the non-significant effects of condition on all main dependent variables, we conducted some exploratory analysis to explore its impact on perceived emotional discomfort and perceived harmfulness. The t-test conducted to compare the means of perceived

harmfulness between participants exposed to hate speech text and those exposed to the neutral text revealed that participants in the exposed condition perceived the text as more harmful (M= 5.70, SD= 1.50) compared to the participants in the neutral condition (M= 4.47, SD= 2.27), $t(172) = -4.23$, $p < .001$. The t-test conducted to compare the means of emotional discomfort between participants exposed to the hate speech text and those exposed to the neutral text showed that participants in the first condition reported higher levels of emotional discomfort (M = 4.03, SD = 1.90) comparing to the latter (M = 2.38, SD = 1.61), $t(172) = -6.20$, $p < .001$.

3.3. Mediation Analysis – Indirect Effects of Exposure to Hate Speech

To test H2, we utilized the SPSS Process macro (Hayes, 2018), model number 4 with bootstrapping with 5,000 samples and 95% bias-corrected confidence intervals. We conducted a two mediation analysis with condition (exposure to hate speech vs. neutral text) as the independent variable, attitudes toward Roma and support for hate speech as the dependent variables, and social norms, perceived harmfulness and emotional discomfort as parallel mediators. Based on the exploratory analysis reported above, we included perceived harmfulness and emotional discomfort as potential mediators. Bivariate correlations confirmed that both variables were significantly linked to our dependent variables of interest (negative attitudes toward Roma and support for hate speech) (see Table 3.1.).

Results showed exposure to hate speech, compared to the neutral condition, triggered more emotional discomfort in participants. In turn, emotional discomfort was negatively related to negative attitudes towards the Roma. The indirect effect of exposure via emotional discomfort was significant for attitudes towards the Roma. (Table 3.2.).

Likewise, participants in the hate speech condition, relative to those in the neutral, showed higher levels of perceived harmfulness in turn, perceived harmfulness was negatively related to negative attitudes towards the Roma. The indirect effect of exposure through perceived harmfulness was also significant for attitudes towards the Roma. (Table 3.2.).

The direct effect of exposure to hate speech (vs neutral) on norms was not significant. Norms were positively associated with negative attitudes toward the Roma, but the indirect effect of condition was not significant (see Table 3.2.)

The results for support for hate speech were very similar. The analysis showed that exposure to hate speech, compared to the neutral condition, led to heightened emotional

discomfort. In turn, this discomfort was negatively associated with support for hate speech, with a significant indirect effect of exposure through emotional discomfort (see Table 3.3.).

Similarly, participants in the hate speech condition, compared to those in the neutral condition, exhibited higher levels of perceived harmfulness. This perception was negatively linked to support for hate speech, and the indirect effect of exposure through perceived harmfulness was also significant for support for hate speech (see Table 3.3.).

The direct effect of exposure to hate speech (vs. neutral) on norms was not significant. Norms were also positively associated with support for hate speech toward the Roma, but the indirect effect of condition was not significant (see Table 3.3.).

Table 3.1.

Means, Standard Deviations, and Correlations of the main dependent variables

Variable	*M*	*SD*	1	2	3	4	5
1. Support for Hate Speech	3.33	1.32	—				
2. Social Norms	5.90	1.50	.29**	—			
3. Perceived Harmfulness	5.10	2.00	−.18*	.11	—		
4. Emotional Discomfort	3.22	1.95	−.20**	.06	.46**	—	
5. Attitudes	4.38	1.51	.76**	.25**	−.28**	−.28**	—

*$p < .05$. **$p < .01$.

Table 3.2.

Indirect Effects of Exposure to Hate Speech on Attitudes Towards the Roma

						R^2
Model 1		Outcome: Social Norms				.004
		Coeff.	SE	*t*	*p*	
	Exposure to Hate Speech	-.19	.23	-.81	.42	
Model 2		Outcome: Perceived Harmfulness				.094
		Coeff.	SE	*t*	*p*	
	Exposure to Hate Speech	1.23	.29	4.23	.000	

Model 3		Outcome: Emotional Discomfort				.183
	Exposure to Hate Speech	Coeff.	SE	*t*	*p*	
		1.66	.27	6.20	.000	
Model 4		Outcome: Attitudes Towards the Roma				.004
		Coeff.	SE	*t*	*p*	
	Exposure to Hate Speech	-.19	.23	-.81	.42	
	Social Norms	.30	.07	4.35	.000	
	Perceived Harmfulness	-.19	.06	-3.19	.002	
	Emotional Discomfort	-.18	.06	-2.89	.004	

Bootsrapping results for indirect effect

	Effect	SE	LL 95% CI	UL 95% CI
Indirect effect of Exposure to Hate Speech on attitudes towards the Roma *via* social norms	-.06	.07	-.20	.08
Indirect effect of Exposure to Hate Speech on attitudes towards the Roma *via* perceived harmfulness	-.23	.08	-.39	-.09
Indirect effect of Exposure to Hate Speech on attitudes towards the Roma *via* emotional discomfort	-.20	.09	-.40	-.04

Unstandardized regression coefficients are reported. 5000 bootstrap samples; LL – lower limit; UL – upper limit; CI – Confident interval

Table 3.3.

Indirect Effects of Exposure to Hate Speech on Support for Hate Speech

		R^2
Model 1	Outcome: Social Norms	.004

		Coeff.	SE	*t*	*p*	
	Exposure to Hate Speech	-.19	.23	-.81	.42	
Model 2	Outcome: Perceived Harmfulness					.094
		Coeff.	SE	*t*	*p*	
	Exposure to Hate Speech	1.23	.29	4.23	.000	
Model 3	Outcome: Emotional Discomfort					.183
	Exposure to Hate Speech	Coeff.	SE	*t*	*p*	
		1.66	.27	6.20	.000	
Model 4	Outcome: Support for Hate Speech					.179
		Coeff.	SE	*t*	*p*	
	Exposure to Hate Speech	.51	.21	2.46	.02	
	Social Norms	.30	.06	4.77	.000	
	Perceived Harmfulness	-.12	.05	-2.24	.03	
	Emotional Discomfort	-.15	.06	-2.63	.01	

Bootsrapping results for indirect effect

	Effect	SE	LL 95% CI	UL 95% CI
Indirect effect of Exposure to Hate Speech on support for hate speech *via* social norms	-.06	.07	-.19	.09
Indirect effect of Exposure to Hate Speech on support for hate speech *via* perceived harmfulness	-.14	.07	-.28	-.02
Indirect effect of Exposure to Hate Speech on support for hate speech *via* emotional discomfort	-.25	.12	-.50	-.03

Unstandardized regression coefficients are reported. 5000 bootstrap samples; LL – lower limit; UL – upper limit; CI – Confident interval

CHAPTER 4

Discussion

The primary goal of this study was to examine how hate speech by state representatives affects attitudes toward the Roma community in Bulgaria. The study aimed to determine whether exposure to hate speech would increase support for hate speech, reinforce negative attitudes, and normalize discriminatory social norms toward the Roma. The study initially aimed to explore if social norms played a role in mediating these effects.

The results indicated that exposure to hate speech did not significantly influence participants' attitudes toward the Roma community, their support for hate speech and for discriminatory social norms which contradicts hypotheses. These findings are inconsistent with previous research by Soral et al. (2018), which suggested that repeated exposure to hate speech increases prejudice against the targeted groups. The findings are also not in line with the previous research, such as Müller and Shwarz (2020), who found that exposure to political hate speech correlates with increased support for discriminatory behavior, and research showing that expressions of hate can lead to increased stereotyping of marginalized communities (Collins & Clément, 2012). The small sample and the context in which the study was conducted could be one possible explanation for this discrepancy; pre-existing neutral or positive attitudes toward the Roma may have mitigated the impact of the hate speech message. Additionally, we can speculate that these findings are related to the hate speech material that might have been perceived as too moderate to provoke substantial shifts in opinions; also, not to forget that the initial low reliability of the hate speech support measure might have impacted the detection of significant differences.

Furthermore, the mediation analysis aligns with previous research by Álvarez-Benjumea and Winter (2020). Our mediation analysis revealed that emotional discomfort and perceived harmfulness mediated the effect of exposure to hate speech on attitudes toward the Roma, which aligns with their research in which the role of emotional reactions in processing and responding to hate speech is emphasized. Contrary to what we expected, social norms did not serve as a significant mediator in these relationships. One potential reason for that could be the prevalence of hate speech directed at the Roma community in Bulgaria, which is deeply rooted and resistant to alterations through experimental manipulations. People in Bulgaria may already hold seated beliefs towards the Roma in Bulgaria that are hard to change.

Another explanation could be the normalized and usual high occurrence of hate speech targeted at the Roma, which in turn cannot be easily influenced by experiments. Suggesting that while emotional factors played a role in mediating the effects of hate speech, social norms need further research.

Despite the added value from these findings, there are still some limitations to consider. To begin with, the research may not be readily applicable to an audience due to the sample size used in the study. Although based on actual political statements, the manipulation of hate speech may not have been strong enough to provoke significant changes in attitudes or behavior. Lastly, relying on self-reported evaluations raises the possibility of social desirability bias in hate speech and discrimination. In future research, these findings should be replicated with a more diverse sample to enhance the generalizability of the results. Different and stronger manipulations of hate speech could be used to better capture the effects on attitudes and behaviors. As mentioned before, social norms should be analyzed further, as no effect was shown in this research. Finally, examining some variables, such as political orientation and prior contact with minority groups, as moderators could be interesting.

References

Álvarez-Benjumea, A., & Winter, F. (2020). The breakdown of antiracist norms: A natural experiment on hate speech after terrorist attacks. *Proceedings of the National Academy of Sciences, 117*(37), 22800-22804. https://doi.org/10.1073/pnas.2007977117

Anderson, K. F. (2013). Diagnosing discrimination: Stress from perceived racism and the mental and physical health effects. *Sociological Inquiry, 83*(1), 55-81. https://doi.org/10.1111/j.1475-682X.2012.00433.x

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology, 41*, 3-33. https://doi.org/10.1111/pops.12670

Bleich, E. (2011). The rise of hate speech and hate crime laws in liberal democracies. *Journal of Ethnic and Migration Studies, 37*(6), 917-934. https://doi.org/10.1080/1369183X.2011.576195

Brauer, M., & Chaurand, N. (2010). Descriptive norms, prescriptive norms, and social control: An intercultural comparison of people's reactions to uncivil behaviors. *European Journal of Social Psychology, 40*(3), 490–499. https://doi.org/10.1002/ejsp.640

Brown, A. (2017). What is hate speech? Part 2: Family resemblances. *Law and Philosophy, 36*, 561-613. https://doi.org/10.1007/s10982-017-9300-x

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity, and compliance. *The Handbook of Social Psychology*, 151–192. http://www.communicationcache.com/uploads/1/0/8/8/10887248/social_influence_-_social_norms_conformity_and_compliance_1998.pdf

Cohen-Almagor, R. (2011). Fighting hate and bigotry on the Internet. *Policy & Internet, 3*(3), 1-26. https://doi.org/10.2202/1944-2866.1059

Collins, K. A., & Clément, R. (2012). Language and prejudice: Direct and moderated effects. *Journal of Language and Social Psychology, 31*(4), 376–396. https://doi.org/10.1177/0261927X12446611

Cortés, I. (2021). Hate speech, symbolic violence, and racial discrimination. Antigypsyism: what responses for the next decade?. *Social Sciences, 10*(10), 1-13. https://doi.org/10.3390/socsci10100360

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: the struggle for internalization. *Journal of Personality and Social Psychology, 82*(3), 359-378. https://doi.org/10.1037//0022-3514.82.3.359

Dimitrova, R., Chasiotis, A., Bender, M., & van de Vijver, F. J. R. (2013). Collective Identity and Well-Being of Bulgarian Roma Adolescents and Their Mothers. *Journal of Youth and Adolescence, 43*(3), 375–386. https://doi.org/10.1007/s10964-013-0043-1

Downs, D. M., & Cowan, G. (2012). Predicting the importance of freedom of speech and the perceived harm of hate speech. *Journal of Applied Social Psychology, 42*(1), 1–23. http://dx.doi.org/10.1111/j.1559-1816.2012.00902.x

Fasoli, F., Paladino, M. P., Carnaghi, A., Jetten, J., Bastian, B., & Bain, P. G. (2016). Not “just words”: Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. *European Journal of Social Psychology, 46*(2), 237-248. https://eprints.qut.edu.au/90602/11/EJSP_dehumanization_uncorrected.pdf

Gelber, K., & McNamara, L. (2015). Evidencing the harms of hate speech. *Social Identities*, *22*(3), 324–341. http://dx.doi.org/10.1080/13504630.2015.1128810

Gelber, K. (2019). Differentiating hate speech: a systemic discrimination approach. *Critical Review of International Social and Political Philosophy*, *24*(4), 393–414. https://doi.org/10.1080/13698230.2019.1576006

Giroud, A., Visintin, E. P., Green, E. G., & Durrheim, K. (2021). 'I don't feel insulted': Constructions of prejudice and identity performance among Roma in Bulgaria. *Journal of Community & Applied Social Psychology, 31*(4), 396-409. https://doi.org/10.1002/casp.2524

Gorenc, N. (2022). Hate speech or free speech: an ethical dilemma? *International Review of Sociology, 32*(3), 413-425. https://doi.org/10.1080/03906701.2022.2133406

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Kende, A., & Krekó, P. (2020). Xenophobia, prejudice, and right-wing populism in East-Central Europe. *Current Opinion in Behavioral Sciences, 34*, 29-33. https://doi.org/10.1016/j.cobeha.2019.11.011

Kende, A., Hadarics, M., & Lášticová, B. (2017). Anti-Roma attitudes as expressions of dominant social norms in Eastern Europe. *International Journal of Intercultural Relations, 60*, 12-27. https://doi.org/10.1016/j.ijintrel.2017.06.002

Kenny, D. A. (2017, February). *MedPower: An interactive tool for the estimation of power in tests of mediation* [Computer software]. Available from https://davidakenny.shinyapps.io/MedPower/.

Langton, R. (1993). Speech acts and unspeakable acts. *Philosophy & Public Affairs*, *22*(4), 293-330. https://www.jstor.org/stable/2265469

Matsuda, M. J. (2018). Public response to racist speech: Considering the victim's story. In *Words that wound* (pp. 17-51). Routledge. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/250f0b14-3300-4772-8fbd-ccf89eb9b9b2/content

Müller, K., & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association, 19*(4), 2131-2167. https://doi.org/10.1093/jeea/jvaa045

Nadal, K. L., Griffin, K. E., Wong, Y., Hamit, S., & Rasmus, M. (2014). The impact of racial microaggressions on mental health: Counseling implications for clients of color. *Journal of Counseling & Development, 92*(1), 57-66. https://doi.org/10.1002/j.1556-6676.2014.00130.x

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology, 96*(3), 574–587. http://dx.doi.org/10.1037/a0011989.supp

Papcunová, J., Martončik, M., Fedáková, D., Kentoš, M., Bozogáňová, M., & Adamkovic, M. (2021). Hate speech conceptualization: A preliminary examination of hate speech indicators and structure. *Complex & Intelligent Systems, 9*, 2827-2842. https://scholar.google.pt/scholar?hl=en&as_sdt=0%2C5&as_vis=1&q=+doi+Hate+speech+conceptualization%3A+A+preliminary+examination+of+hate+speech+indicators+and+structure&btnG=

Parekh, B. (2012). Is There a Case for Banning Hate Speech? In M. Herz & P. Molnar (Eds.), *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (pp. 37-56). Cambridge University Press. https://doi.org/10.1017/CBO9781139042871.006

Pogány, I. (2006). Minority rights and the Roma of Central and Eastern Europe. *Human Rights Law Review, 6*(1), 1-25. https://doi.org/10.1093/hrlr/ngi034

Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior, 44*(2), 136-146. https://doi.org/10.1002/ab.21737

Symvoulakis E. K., Volkos P., Markaki A., et al. (2022). Emotional Discomfort Scale: Instrument Development and Association With General Self-Efficacy and Data From an Urban Primary Care Setting. *Cureus, 14*(1), e21495. https://doi.org/10.7759/cureus.21495

Tankard, M. E., & Paluck, E. L. (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review, 10*(1), 181-211. https://doi.org/10.1111/sipr.12022

Van Baar, H. (2011). Europe's Romaphobia: problematization, securitization, nomadization. *Environment and Planning D: Society and Space, 29*(2), 203-212. https://doi.org/10.1068/d2902ed1

Visintin, E. P., Green, E. G. T., Falomir-Pichastor, J. M., & Berent, J. (2019). Intergroup contact moderates the influence of social norms on prejudice. *Group Processes & Intergroup Relations, 1*(1), 1–23. https://doi.org/10.1177/1368430219839485

Wachs, S., Wettstein, A., Bilz, L., Krause, N., Ballaschk, C., Kansok-Dusche, J., & Wright, M. F. (2022). Playing by the rules? An investigation of the relationship between social norms and adolescents' hate speech perpetration in schools. *Journal of Interpersonal Violence, 37*, 21-22. https://doi.org/10.1177/08862605211056032

White, M. H. II, & Crandall, C. S. (2017). Freedom of racist speech: Ego and expressive threats. *Journal of Personality and Social Psychology, 113*(3), 413–429. https://doi.org/10.1037/pspi0000095

Appendix A – Informed Consent

Информирано съгласие

Целта на това изследване е да разберем как хората се чувстват спрямо речите на обществени фигури и какво мислят за тях. Ще ви помолим да прочетете кратък текст и да отговорите на няколко въпроса.

Изследването се провежда от Мария Абу Ейд-Тотева (Maria_Toteva@iscte-iul.pt) и е научно ръководено от д-р Рита Герра (ana_rita_guerra@iscte-iul.pt), с които можете да се свържете за изясняване на въпроси или споделяне на коментари.

Вашето участие в изследването се цени високо, тъй като ще допринесе за напредъка на знанието в тази област на науката. То се състои в прочитането на кратка реч на обществена фигура и след това отговаряне на няколко въпроса. Очаква се изследването да отнеме по-малко от 20 минути.

Не се очакват значителни рискове, свързани с участието в изследването, но някои от речите могат да включват съдържание, което потенциално може да причини дискомфорт на някои участници. Ако това се случи, можете да се свържете с изследователския екип.

Участието в изследването е строго **доброволно:** можете свободно да изберете дали да участвате или не. Ако сте решили да участвате, можете да прекратите участието си по всяко време, без да е необходимо да предоставяте обосновка. Освен че е доброволно, вашето участие е също анонимно и конфиденциално. Получените данни са предназначени само за статистическа обработка и нито един от отговорите няма да бъде анализиран или съобщен индивидуално. В нито един момент от изследването няма да ви бъде поискано да се идентифицирате.

Съгласно правилата на Iscte, оригиналните данни, събрани за това проучване, ще бъдат **унищожени 6 месеца** след защитата на дисертацията.

Декларирам, съм разбрал/а целите на това, което ми беше предложено и обяснено от изследователя, че имах възможността да задам въпроси относно това изследване, че получих обяснение на всички такива въпроси и че приемам да участвам в изследването.

- ○ Приемам да участвам (1)
- ○ Не приемам да участвам (4)

Appendix B – Demographics

Q2 На колко години сте? (Моля, посочете възрастта си с число)

_____

Q3 Моля, отбележете пола си:

- ○ Жена (1)
- ○ Мъж (2)
- ○ Определям пола си като (моля, посочете) (3)
  _____
- ○ Предпочитам да не отговарям на този въпрос (4)

Q18 Гражданин ли сте на България?

- ○ Да (1)
- ○ Ако не, моля посочете страната, на която сте гражданин (2)
  _____

Q4 Моля, изберете най-високата степен на образование, която сте завършили:

- ○ Начално училище (1)
- ○ Средно училище (2)
- ○ Бакалавърска степен (3)

- ○ Магистърска степен (4)
- ○ Докторска степен (5)
- ○ Избирам да не посоча (6)

Q5 Кое от предложените описания отразява най-точно личната Ви преценка за Вашите доходи?

- ○ Живея комфортно с доходите, които имам. (1)
- ○ Доходите, които имам, ми дават възможност да се справям в живота. (2)
- ○ Доходите, които имам, правят живота ми труден. (3)
- ○ Доходите, които имам, правят живота ми извънредно труден. (4)
- ○ Не знам (5)

Appendix C – Absence of Hate Speech Condition

Q8 Роден през 1968 година, Георги Господинов, изтъкнат романист и поет, е най-признатият съвременен български автор в световен мащаб, като неговите литературни творби преминават границите чрез преводи на 25 езика. Господинов влезе в историята като първият български писател, който спечели престижната награда Букър — връх в литературния свят.

Неговият роман „Времеубежище“ получи тази престижна награда на церемония в Лондон, която отличава изключителни произведения на художествена литература, преведени на английски от цял свят. Изразявайки благодарност за това забележително постижение, на 24 май кметът на София похвали наградата Букър като монументален триумф за българския език и реч. Признавайки дарбата на Господинов, той очаква значително увеличение на читателите, вярвайки, че триумфът ще остави дълготраен отпечатък.

Министърът на културата, подчертавайки деликатната природа на културата, изтъкна нейната стойност и лекотата, с която може да бъде изгубена. Той подчерта важността на отглеждането и призова младите хора да се потопят в света на книгите. Той и кметът насърчиха младежта да чете повече, като развиват отворен подход към литературата в нашия уникален и достъпен глобален пейзаж.

Appendix D – Hate Speech Condition

Q7 В началото на 2019 г., бивш високопоставен държавен служител в България призова за създаването на обширен план за справяне с така наречения „цигански въпрос“. Моля, прочетете кратък откъс от неговата реч. После ще ви зададем няколко въпроса върху текста.

"Циганите в България са станали изключително нагли. Преди няколко години те нападнаха полицай, а преди два дни – военнослужещ. Те са диви зверове, а толерантността на българското общество намалява. Къщите им не отговарят на законовите и нормативните стандарти; наложително е да се извършат щателни проверки на всички гета и да бъдат съборени. Не бива да се очаква от хората да толерират група, която претендира само за права, без да признава задължението си да спазва закона. Българският народ повече няма да подкрепя социалните помощи за циганите. Социални помощи трябва да се дават само на хора, които получават образование и работят, за разлика от повечето цигани в тази страна. Българският народ няма да толерира насилието, и ние ще дадем отпор."

Appendix E – Social Norms

Q11 Моля, посочете степента, в която се съгласявате със следните твърдения.

	1 напълно несъгласен (1)	2 (2)	3 (4)	4 (5)	5 (6)	6 (7)	7 напълно съгласен (8)
Българите изпитват негативни чувства към ромите (1)	○	○	○	○	○	○	○
В България е приемливо да се изразяват негативни чувства към ромите (2)	○	○	○	○	○	○	○

Appendix F – Attitudes

Q10 Моля, посочете степента, в която се съгласявате със следните твърдения

	1 напълно несъгласен (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 напълно съгласен (8)
Ромите са склонни да извършват повече криминални деяния от другите хора (1)	○	○	○	○	○	○	○
Има много малко роми с правилно поведение или разумни роми (2)	○	○	○	○	○	○	○
Ромите нямат положително отношение към работата, те са мързеливи (3)	○	○	○	○	○	○	○
Нарастващото ромско население застрашава сигурността на обществото (4)	○	○	○	○	○	○	○
Ромите обикновено имат много деца, за които не полагат достатъчно грижи (5)	○	○	○	○	○	○	○

Правилно е, че все още има клубове, в които не се допускат роми (6)	○	○	○	○	○	○	○
Смятам, че ромите в тази страна са третирани преференциално в някои аспекти (7)	○	○	○	○	○	○	○
На ромите се дават повече държавни пари, отколкото би трябвало да им се дават (8)	○	○	○	○	○	○	○
Истинската вреда се причинява от организации, които предлагат незаслужено предимство на ромите (9)	○	○	○	○	○	○	○

Appendix G – Hate Speech

Q9 Моля, посочете степента, в която се съгласявате със следните твърдения.

	1 напълно несъгласен (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 напълно съгласен (7)
Съгласен/ съгласна съм с възгледите и ценностите на ромите (1)	○	○	○	○	○	○	○
Ромите са склонни да участват в нечестни практики (2)	○	○	○	○	○	○	○
Ромите са мързеливи и не заслужават доверие (3)	○	○	○	○	○	○	○
Ромите са като болест за нашето общество (4)	○	○	○	○	○	○	○
Ромите трябва да бъдат спрени физически (5)	○	○	○	○	○	○	○

Appendix H – Manipulation Check (Intensity of Hate Speech)

Q12 До каква степен смятате, че текстът, който прочетохте, съдържа реч на омразата?

- 1 Изобщо (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 Много (7)

Appendix I – Manipulation Check (Targeted Group Impact)

Q14 Колко негативно, според вас, влияе мнението на публичните личности на отношението към ромската общност?

- ○ 1 Изобщо не влияе негативно (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (5)
- ○ 6 (6)
- ○ 7 Изключително негативно (7)

Appendix J – Manipulation Check (Perceived Harmfulness)

Q15 Колко вредна смятате, че е речта по отношение на насърчаването на дискриминация или насилие?

- ○ 1 Изобщо не е вредна (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (5)
- ○ 6 (6)
- ○ 7 Изключително вредна (7)

Appendix K – Manipulation Check (Emotional Discomfort)

Q16 До каква степен текстът ви накара да се почувствате емоционално некомфортно?

- ○ 1 Изобщо не се почувствах некомфортно (1)
- ○ 2 (2)
- ○ 3 (3)
- ○ 4 (4)
- ○ 5 (5)
- ○ 6 (6)
- ○ 7 Почувствах се изключително некомфортно (7)

Appendix L – Demographics II

Q6 Като цяло, се смятам за...

- Силно либерален 1 (1)
- 2 (2)
- 3 (3)
- Неутрален 4 (4)
- 5 (5)
- 6 (6)
- Силно консервативен 7 (7)

Q20 До каква степен се идентифицирате с другите българи?

- 1 Напълно не се идентифицирам (1)
- 2 (2)
- 3 (3)
- 4 Неутрален (4)
- 5 (5)
- 6 (6)
- 7 Напълно се идентифицирам (7)

Q21 Моля, посочете доколко сте съгласни с всяко едно от следните твърдения

	Напълно несъгласен 1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	Напълно съгласен 7 (7)
Българите заслужават специално отношение (1)	○	○	○	○	○	○	○
Никога няма да съм доволен/доволна, докато българите не получат всичко, което заслужават (2)	○	○	○	○	○	○	○
Дразня се, когато други критикуват българите (3)	○	○	○	○	○	○	○
Ако българите имаха решаваща роля в света, той щеше да е много по-добро място (4)	○	○	○	○	○	○	○
Изглежда, че малко хора разбират напълно важността на българите (5)	○	○	○	○	○	○	○

Appendix M – Debriefing

Обяснение за изследването

Уважаеми участници, благодарим ви за ценния принос към нашето изследване. Както беше посочено в началото на вашето участие, ние се интересувахме от това как хората се чувстват относно речите на обществените фигури. Специално, това изследване целеше да изследва въздействието на речите на политическите фигури, които съдържат омразна реч към ромите в България.

За целта използвахме обща процедура в психологическите изследвания, където някои участници случайно бяха разпределени да видят различна информация. В този случай някои участници прочетоха измислен текст с омразна реч, а други - неутрален текст. Текстовете, използвани в това проучване, въпреки че са базирани на съществуващи изследвания и медийни представяния, не са реални и са създадени от изследователския екип само за изследователски цели. Основната цел беше да се изследва дали хората, които са изложени на речи на обществени фигури, съдържащи омразни послания към ромите, изразяват по-негативни отношения и дискриминация към тази общност. Важно е да се подчертае, че омразната реч има много негативни последици и няколко проучвания показват, че излагането на омразна реч може да увеличи стереотипите, социалната дистанция от малцинството и подкрепата за дискриминационни политики (Winiewsky и др., 2016). Тя може да насърчи обезчовечаването (Fasoli и др., 2016), недоверието в професионалисти, принадлежащи към стигматизирана група (напр. хора, които имат анти-чернокожи вярвания, не биха наели чернокож адвокат) (Greenberg и Pyszczynski, 1985) и повтарящото се излагане на омразна реч може да повлияе негативно на чувствителността към нея (Soral и др., 2018). Ако след провеждане на това проучване усетите някой от тези ефекти или изпитате стрес, моля използвайте посочените по-долу контакти, за да получите подкрепа.

Ако сами сте жертва на дискриминация, познавате някого, който може да се нуждае от подкрепа или искате да научите повече за расовата дискриминация? Консултирайте https://www.ombudsman.bg/bg Ако искате да научите повече за вредните последици от омразната реч и как да се противопоставите на нея: https://unesdoc.unesco.org https://www.undocs.org

Научното ни изследване се придържа към строги етични насоки, което гарантира поверителност, анонимност и информирано съгласие, както и правото на оттегляне от участие по всяко време. Разбираме чувствителността на темата и сме взели всички необходими предпазни мерки за обработка на данните с най-голямо уважение и конфиденциалност. Ако имате въпроси, коментари или желаете да научите повече за резултатите от изследването, моля, не се колебайте да се свържете с нас. Вашите виждания и обратна връзка са изключително ценни за успеха на нашето проучване и за по-широкото разбиране на въздействието на омразната реч в обществото.

Ще направим основните резултати и заключения от проучването достъпни за всички заинтересовани участници. Ако желаете да получите тази информация, моля посочете

интереса си, и ние ще се погрижим да бъдете информирани за резултатите от изследването. Отново благодарим за вашето участие и принос към това важно изследване.