Exposure to hate speech increases prejudice through desensitization

Wiktor Soral¹ | Michał Bilewicz² | Mikołaj Winiewski²

¹ Institute for Social Studies, University of Warsaw, Warsaw, Poland
² Faculty of Psychology, University of Warsaw, Warsaw, Poland

Correspondence
Wiktor Soral, Institute for Social Studies, University of Warsaw, Stawki 5/7, 00-183 Warsaw, Poland.
Email: wiktor.soral@gmail.com

Funding information
Polish Ministry of Science and Higher Education Iuventus Plus, Grant number: IP2014 002273

In three studies (two representative nationwide surveys, N = 1,007, N = 682; and one experimental, N = 76) we explored the effects of exposure to hate speech on outgroup prejudice. Following the General Aggression Model, we suggest that frequent and repetitive exposure to hate speech leads to desensitization to this form of verbal violence and subsequently to lower evaluations of the victims and greater distancing, thus increasing outgroup prejudice. In the first survey study, we found that lower sensitivity to hate speech was a positive mediator of the relationship between frequent exposure to hate speech and outgroup prejudice. In the second study, we obtained a crucial confirmation of these effects. After desensitization training individuals were less sensitive to hate speech and more prejudiced toward hate speech victims than their counterparts in the control condition. In the final study, we replicated several previous effects and additionally found that the effects of exposure to hate speech on prejudice were mediated by a lower sensitivity to hate speech, and not by lower sensitivity to social norms. Altogether, our studies are the first to elucidate the effects of exposure to hate speech on outgroup prejudice.

KEYWORDS
desensitization, hate speech, prejudice

1 | Exposure to hate speech increases prejudice through desensitization

In 2014, the National Football League considered passing a rule penalizing those who use ethnic slurs on the field (King, 2014). Several years earlier, a similar decision had been made by the New York City council (Pilkington, 2007). In 2012, the Council of Europe started a campaign aimed at reducing the presence of hate speech in public spaces. What unites these actions is a desire to protect the rights and well-being of different minority groups. Indeed, previous studies found that public expressions of hate speech affect psychological well-being and the suicide rate among minorities (Mullen & Smyth, 2004), the exclusion of minorities from the society (Mullen & Rice, 2003), the devaluation of minority members (Greenberg & Pyszczynski, 1985), and the discriminatory distribution of public resources (Fasoli, Maass, & Carnaghi, 2014). However, hate speech affects not only the situation of minorities but it also modifies attitudes and opinions of majority group members. Hate speech violates social norms and as such it poses a threat to social order, similar to other forms of intergroup violence. In that respect, most people—particularly normocentric ones—oppose public expressions of hate speech (Bilewicz, Soral, Marchlewsk, & Winiewski, 2017; Crandall, Eshleman, & O’Brien, 2002). In this article, we analyze how the mere presence of hate speech can affect individuals’ beliefs and opinions regardless of people’s initial opposition to it. This will help explain why, despite growing equality, awareness, and political correctness, hate speech can still affect intergroup relations in contemporary societies.

We employ the model of desensitization from aggression research (Carnagey, Anderson, & Bushman, 2007; see also Bartholow, Bushman, & Sestir, 2006; Funk, Baldacci, Pastold, & Baumgardner, 2004; Krahé et al., 2011). This body of research suggests that frequent exposure to aggressive messages, such as those found in video games (Bartholow et al., 2006), movies, and television (Krahé et al., 2011), or
the Internet (Funk et al., 2004), reduces automatic triggering of negative emotional reactions to images, words, or thoughts of violence. We suggest that the desensitization model can also explain how exposure to hate speech can increase prejudiced responses toward the targets of hate speech, that is, by decreasing negative emotional reactions to such forms of verbal aggression. Thus, understanding how hate speech affects attitudes would allow for the design and implementation of more efficient hate speech reduction interventions.

2 | HATE SPEECH AND LISTENERS' PREJUDICE

Psychological studies that demonstrate the impact of hate speech on prejudice remain relatively scarce. In their seminal studies, Greenberg and Pyszczynski (1985) found that the overhearing of an ethnic slur decreased evaluations of a Black debater, however, only in the condition in which he was losing the debate. The authors suggested that the reduced evaluation of the Black target resulted from the activation of negative schemata associated with the ethnic category. Recent findings by Fasoli et al. (2016) lead to similar conclusions, that supraliminal as well as subliminal exposure to homophobic epithets may increase dehumanization and physical distancing from gay men. The authors of these studies argue that homophobic epithets act as markers of deviance associated with dehumanization processes and physical distancing.

These studies addressed the impact of single hate speech incidents on judgments of outgroup members, but they do not explain how chronic exposure to hate speech can affect more general outgroup attitudes. Additionally, it is unclear whether the findings of Greenberg and Pyszczynski (1985) could be transferred to more general forms of prejudice beside the mere activation of stereotype consistent traits. What is more, although Greenberg and Pyszczynski (1985) refer to hate speech as the “disease,” neither they nor Fasoli et al. (2016) point to the exact process by which such a “disease” could spread.

3 | DESENSITIZATION TO HATE SPEECH

To understand the effects of (chronic) exposure to hate speech, one should also understand how individuals react to other unpleasant, threatening or shocking stimuli. Classical studies (Wolpe 1958, 1964) suggest that after initial physiological and (negative) affective arousal individuals gradually learn to ignore such alarming messages and desensitize to further alarms. Importantly, these results were used to understand how individuals habituate to frequent violence and how this process may lead to increased likelihood of aggression (e.g., Cline, Croft, & Courrier, 1973).

The desensitization process is observed at very basic physiological levels. Subsequently displayed violent images are associated with diminishing heart rate (e.g., Linz, Donnerstein, & Adams, 1989) and decreased skin conductance (e.g., Cline et al., 1973). Moreover, violence desensitization is observed at the brain level in the P300 component of the event related potential, associated with the activation of the aversive motivational system (Bartholow et al., 2006).

The General Aggression Model by Carnagey et al. (2007) explains how these physiological responses result in cognitive and affective outcomes that could lead to increased aggressive behavior and a lower likelihood of helping victims. Desensitization processes observed at lower physiological levels (e.g., decreased heart rates and decreased skin conductance) maps the extinction of negative emotional reactions to violence associated with: decreased perception of injury severity (see also Linz et al., 1989), decreased attention to violent events, decreased sympathy for violence victims, increased belief that violence is normative, and decreased negative attitudes toward violence.

We argue that a similar process—the reduction of negative cognitive and affective responses to verbally violent stimuli—is responsible for increased outgroup prejudice after being (chronically) exposed to ethnic slurs and hate speech. Frequent exposure to hate speech would result in lower affective responses to consecutively incoming hostile messages. After passing a certain tipping point, subsequent messages would cease to evoke an affective response. Consequently, hate speech would be interpreted by an individual as less negative and harmful, less important, and less violating of social norms. By decreasing sympathy for the victims of hate speech, and activating stereotypes justifying the use of verbal violence, such exposure would increase prejudice and distancing intentions.

Surprisingly, only one study has examined the effects of previous exposure to hate speech on desensitization: Leets (2001) has shown that participants previously exposed to racial slurs exhibited decreased sensitivity to hate speech. However, this correlational study was performed on a very small sample and had limited internal and external validity. Moreover, the author did not assess how such desensitization could impact attitudes toward outgroups. In order to address these gaps in the literature, we designed three studies examining the role of desensitization in the influence of hate-speech on prejudice.

4 | OVERVIEW OF THE STUDIES

In the present studies we aimed to understand the mechanism by which exposure to hate speech affects outgroup attitudes. We verified the hypothesis that frequent exposure to hate speech would decrease the perception of hate speech harmfulness and that such exposure would subsequently lead to increased outgroup prejudice. We conducted three studies in the Polish context: two correlational studies using a representative nation-wide sample and one experimental study on a small homogenous sample. We aimed to identify the causal mechanisms underlying this process and at the same time we sought to ensure high levels of external validity.
5  STUDY 1

The aim of Study 1 was to explore the relationship between the frequency of a person’s exposure to hate speech and the level of outgroup prejudice. We predicted that frequent contact with hostile statements toward certain minorities would be a risk factor associated with higher levels of prejudice. Moreover, we expected this relationship to be mediated by the level of sensitivity to hate speech. Thus, more frequent contact with hate speech would lead to desensitization and this desensitization would subsequently increase prejudice.

We focused our study on two minorities particularly affected by hate speech in Poland: LGBT and Muslims. Previous studies (Bilewicz, Marchlewksa, Soral, & Winiewski, 2014) have shown that about 50% of adult Poles were exposed in the past to anti-Muslim hate speech. In the case of hate speech targeting LGBT groups this proportion seems to be even higher (up to 73%). At the same time, the level of prejudice toward these two minorities is one of the highest in Poland, and both groups are perceived by Poles as threatening to established values (e.g., Cichocka, Winiewski, Bilewicz, Bukowski, & Jost, 2015). While the sources of this situation are certainly complex, we would like to emphasize one possible explanation: individual level of prejudice toward these two groups stems from the abundance of mass media containing hate speech toward these targets.

5.1  Method

5.1.1  Participants and procedure

To ensure the high level of generalizability, we conducted Study 1 on a representative sample of adult Poles, N = 1,007 adults (524 women, 483 men, M_age = 46.27 years, SD = 17.54), with 91 percent self-declared as religious. The sample was collected using stratified random sampling based on (PESEL) Polish Residence Identification Number held by national census bureau. The final sample was additionally weighted according to several demographic variables in order to match the sample to the most recent Polish census. Data from this sample were collected in face-to-face computer assisted interviews (CAPI).

Participants were presented with a list of preselected examples of hate speech directed against LGBT people and Muslims. After each example, participants were asked to what extent they perceive such statements as offensive versus inoffensive. Afterwards, participants indicated how often they encounter similar statements in their private lives. Finally, each participant completed an outgroup prejudice scale.

5.1.2  Measures

Frequency of exposure to hate speech

We measured the frequency of contact with hate speech using the question “How often do you encounter such statements directed at ...,” addressed at anti-LGBT hate speech (M = 3.53, SD = 2.06) and anti-Muslim hate speech (M = 2.29, SD = 1.67). Participants indicated their responses on a scale from 1 (never) to 7 (very often). Correlation between both items was moderate, r(963) = .48, p < .001, and the reliability of the two-items scale was acceptable, α = .65.

Sensitivity to hate speech

In order to determine individual levels of indifference toward hate speech statements, we used a list of six preselected examples of hostile language. Three of these statements were directed against LGBT group (α = .73, M = 5.55, SD = 1.60, e.g., “I am disgusted by fags, they are a degeneration of humanity and they should be cured”), and three against Muslims (α = .78, M = 5.99, SD = 1.43, e.g., “Muslims are stinky cowards who can only murder women, children and innocent people”). Participants were asked to indicate on a 7-point scale whether they perceived each example as “Not at all offensive” (1) or “Strongly offensive” (7). The full list of statements is included in Supplemental Materials A.

Outgroup prejudice

As a measure of outgroup prejudice we included a Social Distance Scale (Bogardus, 1925, 1933). Participants had to respond whether they would (or not) accept a member of a certain minority as a co-worker, as a neighbor, or as part of their family. Responses ranged from 1 (definitely would accept) to 4 (definitely would not accept). By averaging the scores across these questions we obtained a single measure of prejudice toward LGBT (α = .82, M = 2.28, SD = 0.90) and Muslims (α = .82, M = 2.12, SD = 0.85).

5.2  Results

5.2.1  Analytical strategy

Data were analyzed in line with the structural equation models approach. The reported frequency of exposure to hate speech directed against LGBT and Muslims formed a latent variable of general exposure to hate speech. Reported sensitivity to hate speech directed against LGBT and Muslims formed two lower level latent variables, which subsequently formed a single latent variable associated with general sensitivity to hate speech. Similarly, reported social distance formed two lower level latent variables indicating the level of prejudice toward LGBT and Muslims, and these in turn formed single latent variable of participants’ general level of prejudice. In order to account for the similarity of the item’s phrasing, constraints on error correlations between items concerning workplace, neighborhood, and family were released. Thus, the measurement part of the model formed a hierarchy aimed at revealing a general structure of attitudes. For zero-order correlations between latent variables, see Table 1 (also see zero-order correlations between observed variables in Supplemental Materials B).

In the structural part of the model, we first regressed general sensitivity to hate speech on frequency of exposure to hostile statements. In a second equation, we regressed outgroup prejudice on sensitivity to hate speech and the frequency of exposure to hate...
sensitize people to hate speech, and make them more attentive to hate speech. This allowed us to obtain an indirect effect of exposure to hate speech on outgroup prejudice via sensitivity to hate speech. The model was fitted with maximum likelihood estimator with bootstrapped standard errors (B = 1,000). To manage missing responses, we performed a full information maximum likelihood estimation.

5.2.2 | Effects of exposure to hate speech on sensitivity to hate speech and outgroup prejudice

The model with estimated coefficients is presented on Figure 1. The model obtained a good fit (for specific indices see caption of Figure 1). Exposure to hate speech was a significant and negative predictor of sensitivity to hate speech.

Subsequently lower levels of sensitivity to hate speech were associated with stronger outgroup prejudice. Moreover, a nonsignificant total effect of exposure to hate speech on outgroup prejudice, after accounting for the effect of sensitivity to hate speech, became a significant negative predictor of outgroup prejudice, which suggests a suppressive effect of sensitivity to hate speech. The magnitude of change (treated equivalently as a size of indirect effect) was statistically significant.

5.3 | Discussion

The study found that people who frequently encounter examples of hate speech are less inclined to perceive hate speech as an offensive and abusive phenomenon. This desensitization to the harmfulness of hate speech was in turn a risk factor of greater outgroup prejudice (measured with Social Distance Scale). However, we did not observe any evidence of a simple relation between the frequency of exposure to hate speech and prejudice.

Surprisingly, after removing a part of the variance associated with the perception of hate speech offensiveness, the direct effect of exposure on distance was negative. That is, keeping other factors constant, those more exposed to hate speech were preferring lower distance toward outgroup members. One possible explanation might be related to the fact that people less distanced toward outgroups (LGBT people and Muslims), for example, those having friends among members of the outgroup, might be more likely to recognize the cases of hostile language use. Positive outgroup attitudes might as well sensitize people to hate speech, and make them more attentive to hate speech in their environment. Another possible explanation is that contact with hate speech might also have some positive consequences (e.g., raised compassion for the victims) on intergroup relations. However, such positive consequences are suppressed by the desensitization process. The correlational nature of this study means that some other possible interpretations of this relationships cannot be excluded. For instance, awareness of social desirability norms and motivation to control prejudice (Dunton & Fazio, 1997) can also produce sensitivity that can be responsible for similar outcomes. Therefore, the correlational character of this study did not allow us to assess the exact nature of the desensitization process. This is why we designed a second, experimental, study that would allow us to establish the causal path implied by the hate speech desensitization hypothesis.

6 | STUDY 2

In Study 2 we aimed to establish a causal link between exposure to hate speech and decreased sensitivity to hate speech utterances, which subsequently would lead to increased outgroup prejudice. Numerous studies point to the prevalence of hate speech in public space. In such an environment, individuals can be confronted with hostile language for a very long period of time. However, we examined whether relatively short but systematic exposure to hate speech would have an effect on intergroup attitudes. We predicted that even such short exposure would lead to perceiving hate speech as being less harmful and to becoming more prejudiced toward the targets of hate speech.

6.1 | Method

6.1.1 | Participants and design

Seventy-six undergraduate students (all Polish and White) took part in a study. The sample size was calculated with an aim to ensure adequate power (.80) to detect medium-size effects (with
one-sided \( p = .05 \) and established at 40 persons per condition. The sample was composed of 63 female and 12 male participants with a mean of age of 23.21 years (SD = 3.95). Participants were randomly assigned to one of two conditions: control (\( n = 36 \)) or experimental (hate speech, \( n = 39 \)). One person resigned from taking part in the study in the middle of the procedure, thus leaving a total of 75 participants.

### 6.1.2 Procedure and measures

After arriving at the laboratory and signing a consent form, participants were told that they will take part in a study on the relationship between web design and memory processes. This part constituted the desensitization training and the participants’ task was to read through five web-pages from discussion forums and to assess the esthetics and readability of the page design. After each page, the participants answered a single control question in order to check whether they had read the content of the page carefully. Each participant was informed that at the end of the study they would take part in a test aimed at checking how well they memorized the content of the pages. In the control condition, the content consisted of neutral comments selected from Internet discussion forums. In the experimental condition, each page consisted of examples of hate speech directed against one of five different minority group: Jews, Ukrainians, Roma, LGBT, and Muslims. The full list of comments is included in Supplemental Materials A.

#### Sensitivity to hate speech

To measure the perceived offensiveness of hate speech, we used a list of 15 statements that had been pre-selected from the Internet. The statements were different from those used in the experimental condition of the desensitization training and were directed against five minorities (three statements directed against the same five targets as in the desensitization training). Participants had to indicate on a 7-point whether they perceived each example as not at all offensive (1) vs strongly offensive (7). We created one sensitivity to hate speech indicator by averaging the scores on the 15 statements in the desensitization training (\( \alpha = .89, M = 6.17, SD = 0.76 \)).
Outgroup prejudice

We measured outgroup prejudice toward groups presented in the desensitization training. As in Study 1, we used a modified version of the Social Distance Scale (Bogardus, 1925, 1933). Participants had to indicate on a 4-point scale whether they would accept (1) or not (4) a member of a certain group as a co-worker, a neighbor, or a part of their family. We aggregated responses for groups presented in the desensitization training ($\alpha = .85, M = 1.56, SD = 0.42$).

Manipulation check procedure

We measured the perceived hostility of comments in the desensitization training with a single question: To what extent were comments in the first part of the study hostile? Participants indicated their responses on a scale ranging from 1 (not at all) to 4 (very much).

6.2 Results

6.2.1 Initial analyses

To verify the efficiency of the manipulation, we compared the perceived hostility of comments in the two conditions. The analysis confirmed our expectations in that the comments in the experimental version of the training were perceived as being significantly more hostile than in the control condition ($M_{c} = 3.19$, $SD = 0.62$ vs. $M_{x} = 3.77$, $SD = 0.58$); Yuen–Welch test with $B = 5,000$ samples: $t(73) = 4.26$, $p < .001$, $d = 1.46$.

6.2.2 Effects of desensitization training on attitudes toward groups targeted by hate speech

To check whether exposure to hate speech would lead to desensitization and subsequently to increased outgroup prejudice toward groups targeted by hostile language, we performed a mediation analysis. First, we regressed sensitivity to hate speech on a dummy-coded variable indicating one of two conditions (0—control condition, 1—experimental condition). In a second equation, we regressed outgroup prejudice on the sensitivity to hate speech, while controlling the effect of the experimental condition. The obtained model is presented in Figure 2.

As one can see in Figure 2, participants in the experimental condition ($n = 39$) perceived examples of hate speech as being significantly less offensive than participants in the control ($n = 36$) condition ($M_{c} = 6.40$, $SD = 0.53$ vs. $M_{x} = 5.95$, $SD = 0.87$), Yuen–Welch test with $B = 5,000$ samples: $t(73) = 2.30$, $p = .023$, $d = 0.54$. At the same time participants in the experimental condition (exposed to hate speech) exhibited significantly higher levels of prejudice than participants in the control condition ($M_{c} = 1.46$, $SD = 0.37$ vs. $M_{x} = 1.65$, $SD = 0.44$), Yuen–Welch test with $B = 5,000$ samples: $t(73) = 1.98$, $p = .048$, $d = 0.44$. However, after controlling for the level of sensitivity to hate speech, the estimated difference reduced to nonsignificant, while the level of sensitivity to hate speech significantly predicted outgroup prejudice. Estimated average causal mediation effect (i.e., indirect effect, see Tingley, Yamamoto, Hirose, Keele, & Imai, 2014) was positive and significant, $ACME = 0.11$ (95% quasi-bayesian CI: 0.03, 0.20). Additional sensitivity analysis indicated a moderate degree of robustness of the finding to the effects of non-observed pre-treatment covariates (see, Imai, Keele, & Yamamoto, 2010), guaranteeing ACME positivity as long as a single unobserved covariate explains less than 16% of the residual variance of both desensitization and prejudice.

6.3 Discussion

In the experimental study we demonstrated that even a relatively short exposure to hate speech could desensitize participants to its offensiveness. Moreover, confrontation with hate speech caused an increase in the level of prejudice. In fact, the desensitization process fully mediated this change.

7 STUDY 3

We conducted Study 3 having three different goals in mind. First, we wanted to replicate the previous findings in the context of the refugee crisis in Europe. Instigated by the war in Syria and other regional turmoils, the extraordinary migration of refugees from the Middle East and Northern Africa to Europe has caused great concern among European citizens. The refugee crisis has dominated the public debate in many European countries and has unfortunately led to an increase in the frequency of hate speech, particularly toward Muslim refugees. Thus, it was important to examine how these forms of hostile language affect attitudes toward refugees, especially among adolescents—the group most often exposed to hate speech online (see Winiewski et al., 2017). Second, we wanted to verify our model with a different type of measure of outgroup prejudice. In previous studies, we used only the Social Distance Scale (Bogardus, 1925, 1933), and it was likely that the results were specific to this particular measure. Therefore, in Study 3 we included an additional measure of anti-immigrant attitudes. Lastly, we wanted to examine an alternative explanation of why exposure to
hate speech affects outgroup prejudice. Recent findings by Hsueh, Yogeeswaran, and Malinen (2015) suggest that exposure to hate may influence social norms which then make it acceptable to express prejudice. In Study 3 we included an additional measure of sensitivity to a general social norm of not harming others as an alternative mediating variable.

7.1 | Method

7.1.1 | Participants and procedure

We conducted Study 3 on a large, nation-wide sample of $N = 682$ adolescent Poles aged from 16 to 18 ($M_{\text{age}} = 16.71$, $SD = 0.62$, 350 women, 332 men), with 84 percent self-declared as religious. Similarly to Study 1, the sample was collected using stratified random sampling based on PESEL database. Data was collected using face-to-face computer assisted interviews (CAPI).

Participants were presented with a list of pre-selected examples of hate speech directed against refugees and Muslims. After each example, participants were asked to what extent they perceive such statements as offensive versus inoffensive. Afterward, participants indicated how often they encounter similar statements in their private lives. Finally, each participant completed a measure of sensitivity to social norms and two different measures of anti-immigrant prejudice.

7.1.2 | Measures

Exposure to hate speech

We used the same measure as in Study 1 to examine the frequency of exposure to hate speech, addressed this time at Muslims ($M = 3.66$, $SD = 2.13$) and refugees ($M = 4.31$, $SD = 2.11$). Participants indicated frequency of encountering hate speech statements on a scale from 1 (Very rarely) to 7 (Very often). Correlation between both items was large, $r (671) = .68$, $p < .001$, and the 2-item scale was reliable, $\alpha = .81$.

Sensitivity to hate speech

As in previous studies, we asked participants to rate the offensiveness of the hate speech comments. Three statements were directed against Muslims (see Study 1, $\alpha = .87$, $M = 5.26$, $SD = 1.67$), and three additional were directed against refugees ($\alpha = .73$, $M = 5.68$, $SD = 1.41$), for example, “Let the refugees arrive to our country. What else will you burn in power plants? The rest of them you can always redo into dog food.” Participants rated the offensiveness of each statement on a scale from 1 (Not at all offensive) to 7 (Strongly offensive). See Supplemental Materials A for a full list of statements.

Sensitivity to social norms

To assess sensitivity to social norms participants were presented with a short story about a taxi driver, who in an act of vengeance harms his competitor in a number of ways (i.e., he blocks his competitor’s car, he covers his competitor’s car with sludge, he damages the brake hoses in his competitor’s car). The examples were preselected to represent actions that are in general perceived to be against social norms. Participants were asked, on a scale from 1 (approved by the majority of the society) to 7 (disapproved by the majority of the society), to assess how much each action would be approved of by the majority of the society. The average of the three responses was treated as an index of sensitivity to social norms ($\alpha = .75$, $M = 5.20$, $SD = 0.89$).

Outgroup prejudice

As previously, we used the Social Distance Scale (Bogardus, 1925, 1933) to measure prejudice toward Muslim refugees ($\alpha = .77$, $M = 2.56$, $SD = 0.87$). Participants indicated whether they would accept (1) or not (4) an outgroup member in their social context.

Anti-immigrant attitudes

As an alternative measure of outgroup prejudice, we have included a 9-item scale of support for different, radical anti-immigrant policies (Winiewski et al., 2017). Participants were asked to indicate on a 7-point scale whether they would support (7) or not (1) a government policy aimed at using physical violence against refugees (three items, $\alpha = .80$, $M = 4.49$, $SD = 1.54$), for example, “Due to the refugee crisis, the government should allow uniformed services to use direct coercion and violence against refugees”; aimed at using psychological violence against refugees (three items, $\alpha = .75$, $M = 4.60$, $SD = 1.47$), for example, “Due to the refugee crisis, the government should allow intelligence services to use surveillance and control the private lives of refugees”; or aimed at excluding refugees from the rest of the society (three items, $\alpha = .87$, $M = 3.78$, $SD = 1.74$), for example, “Refugees arriving in Poland should be settled in special centers isolated from the rest of the society.”

7.2 | Results

7.2.1 | Analytical strategy

Data were analyzed using full SEM approach. Two measures of exposure to hate speech formed a single latent variable of exposure to anti-refugee hate speech. Measures of sensitivity to hate speech directed against Muslims and refugees formed separate latent variables, which then formed a single variable of sensitivity to anti-refugee hate speech. Three measures of social distance and three measures of sensitivity to social norms formed separate latent variables. Finally, 9 measures of anti-immigrant attitudes formed three separate subscales, which then formed one latent variable. All path coefficients in the measurement part of the model were statistically significant, $p < .001$, with standardized coefficients greater than .49. See Table 2 for zero-order correlations between latent variables (also see zero-order correlations between observed variables in Supplemental Materials B).

In the structural part, we first regressed all latent variables: sensitivity to hate speech, sensitivity to social norms, outgroup prejudice, and anti-immigrant attitudes on a latent variable of exposure to hate speech. Additionally, we regressed outgroup prejudice and anti-immigrant attitudes on sensitivity to hate speech and sensitivity to social norms.
7.2.2 The effect of exposure to hate speech on outgroup prejudice and its mediators

The full model obtained a satisfactory fit (see Figure 3 for coefficients, and the figure caption for fit indices). Exposure to hate speech was a significant and positive predictor of both outgroup prejudice and anti-immigrant attitudes. Furthermore, frequent exposure to hate speech was significantly associated with lower sensitivity to hate speech. However, the relation between exposure to hate speech and sensitivity to social norms was not significant. Moreover, lower sensitivity to hate speech was significantly associated with higher prejudice and with increased anti-immigrant attitudes. On the other hand, sensitivity to social norms was not significantly related to prejudice, and its relation with anti-immigrant attitudes was positive but small—those with greater sensitivity to social norms endorsed more anti-immigrant attitudes. After accounting for both mediators, the direct effects of exposure to hate speech on outgroup prejudice and anti-immigrant attitudes became reversed or not significant, respectively.

To further examine the role of mediators, we analyzed estimates of indirect effects with bootstrapped confidence intervals ($B = 5000$). The indirect effects of exposure to hate speech on outgroup prejudice and anti-immigrant attitudes through lower sensitivity to hate speech were both statistically significant, standardized $IE = .25$ (95%CI: .18, .33) and $IE = .22$ (95%CI: .15, .28). The indirect effects of exposure to hate speech on both outgroup prejudice and anti-immigrant attitudes through lower sensitivity to social norms were not significant, standardized $IE = .002$ (95%CI: -.003, .010) and $IE = -.005$ (95%CI: -.016, .002).

7.3 Discussion

In a nation-wide study of Polish adolescents, we aimed to find further evidence for the model of desensitization to hate speech, here in the context of the refugee crisis in Europe. We found that individuals frequently exposed to anti-refugee hate speech were in general more prejudiced toward refugees. This effect was observed not only in the case of rather subtle measures of outgroup prejudice but also it was manifested in greater support for radical, anti-immigrant government policies. Importantly, we found that the effects of exposure to hate speech on prejudice were mediated by a lower level of sensitivity to hate speech. This may suggest that those frequently exposed to hate speech no longer see such statements as offensive, and this results in their lower sympathy for the victims of hate speech and their increased prejudice.

Previous studies suggested that hate speech may influence social norms and this may lead to increased expression of prejudice. In our study, we did not find any support for the role of the social norm of not doing harm to others. We do not claim that social norms are unimportant. On the contrary, according to the General Aggression Model (Carnagey et al., 2007) the increased perception of hate speech as normative behavior is one of the consequences of the desensitization. However, the claim that exposure to hate speech
increases prejudice as a part of a more general change in perception of what is harmful to others is not supported by our results.

8 | GENERAL DISCUSSION

In three studies, including representative sample surveys, and an experimental study, we tested the detrimental effects of exposure to hate speech on intergroup relations. Based on the literature, we predicted that encountering hate speech would lead to greater desensitization to such statements (Leets, 2001) and that it would increase outgroup prejudice expressed toward the victims of those hostile statements (Greenberg & Pyszczynski, 1985). Both of our predictions were confirmed, and moreover we found that the desensitization process seems to be an important link between exposure to hate speech and outgroup prejudice. Thus, exposure to ethnic slurs decreases sympathy for the victims of such verbal violence through the desensitization process (see Carnagey et al., 2007).

In the first study, performed on a large representative sample, we found a significant link between the frequency of exposure to hate speech statements and the degree to which such statements were perceived to be offensive. Exposure to hate speech was indirectly linked to greater outgroup prejudice. However, the direct effect was to reduce prejudice, that is, after controlling for the desensitization, frequent exposure to hate speech was associated with decreased outgroup prejudice. This non-intuitive finding might be due to either the increased sensitivity of non-prejudiced individuals to human right violations or some positive effects of exposure to hate speech on outgroup attitudes. In the second—experimental—study, we were able to resolve this issue. In Study 2 we found that after controlling for the effects of desensitization, the direct link between exposure to hate speech and outgroup prejudice is insignificant, and that desensitization fully mediates the effect. Finally, Study 3, performed on a nation-wide representative sample, showed some of the crucial societal consequences of hate-speech desensitization in the context of immigration policies. This study also allowed us to examine the desensitization mechanism as opposed to changes in norm-violation perceptions.

Obviously, there are certain limitations of the current studies. The finding that desensitization increases outgroup prejudice holds as long as we assume the non-existence of unobservable important pre-treatment covariates of both the level of desensitization and outgroup prejudice. Potential personality variables that could be held responsible for the observed effect are social dominance orientation and right-wing authoritarianism (see, e.g., Bilewicz et al., 2017; Sidanius et al., 2013), thus, future studies should include such variables in the experimental procedure. Another limitation is that our studies focused solely on the reactions to hate speech among majority members. That is, we aimed to assess desensitization processes in the case of minority-directed hate speech. Acknowledging the severe consequences of hate speech for immigrant minorities (Mullen & Smyth, 2004), one could doubt whether being constantly immersed in ingroup-directed hateful language has any desensitizing consequences for minority members. Possibly, in the case of minority members ingroup-directed hate speech could amplify emotional reactions to hate speech rather than to mitigate them. This however needs to be tested in future empirical studies. Finally, in our studies we were able to experimentally induce exposure to hate speech for only a relatively short period of time, and we only measured exposure to hate speech for longer periods. Because our hypotheses concern consequences of the long-term exposure to hate speech, it may be important to address the question of the temporal dynamics of the desensitization processes and its boundaries in future studies. Such new studies should also include more accurate and more objective measures of the frequency of exposure to hate speech—our claims are valid as long as we accept that most of our participants were to some extent accurate in assessing the frequency of their exposure to hate speech.
This research points to the fundamental role of the environment in forming aggression and prejudice on an intergroup level. Chronic exposure to violence is considered one of the key environmental antecedents of aggression according to the General Aggression Model, as it affects both cognitive processes and behavioral outcomes (Allen, Anderson, & Bushman, 2018). We propose that new means of online communication create an environment where verbal aggression is more legitimate. When being frequently exposed to hateful online commentaries, people become increasingly desensitized to them, and ultimately the contents of these commentaries come to shape their perception of outgroup members (minorities, immigrant groups, political adversaries). We found that such desensitization cannot be attributed to changes of norms, but rather to decreased emotional reaction, something that was observed on a very basic physiological level. Without negative emotional reactions to offensive language, people are more willing to believe in its content and to treat its contents as guiding principles (e.g., in case of anti-immigration policies).

Apart from its theoretical importance, this set of studies has obvious applied implications. Facilitating the spread of hate speech in public space makes its recipients less able to notice the harm done by the verbal violence. Ultimately, hate speech affects both the targets of hate speech (e.g., Mullen & Smyth, 2004) and the wider society that witnesses such violence. There is much to be done, especially in the field of hate speech prevention. Thus, further studies should address not only the mechanism of desensitization, but also the factors and conditions that would sensitize individuals to cases of verbal violence.

ACKNOWLEDGMENT

This research was conducted within Polish Ministry of Science and Higher Education Juventus Plus grant (IP2014 002273) to Michał Bilewicz. The data collection was supported by The Stefan Batory Foundation.

CONFLICTS OF INTEREST

All authors declare no conflict of interest.

ENDNOTES

1 The examples were selected from the Wiedza Lokalna Foundation database of online hate speech. The selected examples were presented to representatives of the minorities targeted by hate speech who were asked to assess whether these sentences are considered by them examples of hate speech and offensive language. The examples selected for the study were the ones that were judged by minority members as offensive and hateful.

2 We would like to thank anonymous reviewer for pointing this limitation.

REFERENCES


ORCID

Wiktor Soral http://orcid.org/0000-0003-2557-8071


SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Soral W, Bilewicz M, Winiewski M. Exposure to hate speech increases prejudice through desensitization. Aggr Behav. 2017;1–11. https://doi.org/10.1002/ab.21737