

## RESEARCH ARTICLE

# Direct and indirect effects of social dominance orientation on hate speech perpetration via empathy and moral disengagement among adolescents: A multilevel mediation model

Melisa Castellanos<sup>1</sup> | Alexander Wettstein<sup>1</sup> | Sebastian Wachs<sup>2,3</sup> | Ludwig Bilz<sup>4</sup>

<sup>1</sup>Institute for Research, Development and Evaluation, Bern University of Teacher Education, Bern, Switzerland

<sup>2</sup>Department of Educational Sciences, University of Potsdam, Potsdam, Germany

<sup>3</sup>National Anti-Bullying Research and Resource Centre, Dublin City University, Dublin, Ireland

<sup>4</sup>Department of Health Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany

## Correspondence

Alexander Wettstein, Institute for Research, Development and Evaluation, Bern University of Teacher Education, Room D135, Fabrikstrasse 8, CH-3012 Bern, Switzerland. Email: [alexander.wettstein@phbern.ch](mailto:alexander.wettstein@phbern.ch)

## Funding information

German Research Foundation (DFG), Grant/Award Numbers: BI 1046/9-1, WA 4275/2-1; University of Teacher Education Bern (Pädagogische Hochschule Bern), Grant/Award Number: 19s 0008 01

## Abstract

Although it is known that social dominance orientation directly affects hate speech perpetration, few studies have explored the mechanisms by which this effect takes place during adolescence. Based on the socio-cognitive theory of moral agency, we aimed to fill this gap in the literature by exploring the direct and indirect effects of social dominance orientation on hate speech perpetration in offline and online settings. The sample included seventh, eighth, and ninth graders ( $N = 3225$ ) (51.2% girls, 37.2% with an immigrant background) from 36 Swiss and German schools who completed a survey about hate speech, social dominance orientation, empathy, and moral disengagement. A multilevel mediation path model revealed that social dominance orientation had a direct effect on offline and online hate speech perpetration. Moreover, social dominance also had indirect effects via low levels of empathy and high levels of moral disengagement. No gender differences were observed. Our findings are discussed regarding the potential contribution to preventing hate speech during adolescence.

## KEYWORDS

adolescents, cyberhate, empathy, hate speech, moral disengagement, social dominance orientation

## 1 | INTRODUCTION

Hate speech includes derogatory expressions (e.g., words, images, videos) that directly or vicariously target people based on ascribed group characteristics (e.g., ethnicity, nationality, gender, sexual orientation, disability, religion) (Kansok-Dusche et al., 2022). Hate speech can have devastating consequences for victims, perpetrators, and witnesses (Krause et al., 2021; Wachs, Gámez-Guadix, et al., 2022) at the communal and societal level (Kansok-Dusche et al., 2022; Wettstein, 2021). A correlate of hate speech

perpetration is social dominance orientation, which refers to the preference for hierarchies and inequalities among social groups (Pratto et al., 1994; Sidanius & Pratto, 1999). The current literature on hate speech and social dominance orientation is limited in at least three ways: (1) a focus on adult samples, (2) the separate examination of offline and online settings, and (3) a lack of studies about the mechanisms that ease or inhibit the association between social dominance orientation and hate speech perpetration. To overcome these gaps in knowledge, we aimed to investigate: (1) the association between social dominance orientation and adolescents' online and

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Aggressive Behavior* published by Wiley Periodicals LLC.

offline hate speech perpetration and (2) whether there was an indirect effect of social dominance orientation on hate speech perpetration via empathy and moral disengagement, two important correlates of hate speech perpetration during adolescence (Wachs, Bilz, Wettstein, Wright, Kansok-Dusche, et al., 2022). Understanding these underlying mechanisms may help to refine hate speech prevention in adolescence.

## 2 | THEORETICAL BACKGROUND

### 2.1 | Social dominance theory and social dominance orientation

According to the social dominance theory (Pratto et al., 1994, 2006; Sidanius & Pratto, 1999), societies maintain discrimination and prejudice due to the confluence of ideologies, institutional practices, social relationships, and individual characteristics, through ideologies that postulate that some groups are dominant and others are subordinate. Hierarchies are based on three stratification systems: (1) age (i.e., older people have disproportionate power over younger people), (2) gender (i.e., men are better off than women), and (3) the arbitrary system (e.g., race, sexual orientation, or religion; Sidanius & Pratto, 1999, p. 33). Social dominance orientation is defined as the “degree to which individuals desire and support group-based hierarchy and the domination of “inferior” groups by “superior” groups” (Sidanius & Pratto, 1999, p. 48). Social dominance orientation is associated with generalized prejudice (Amiot & Bourhis, 2005; Bäckström & Björklund, 2007; Kteily et al., 2012; McFarland, 2010; Soral et al., 2018), bullying perpetration (Goodboy et al., 2016), and lower probability of banning hate speech (Bilewicz et al., 2017). To the best of our knowledge, no studies exist for youth. In line with the evidence from adult samples, we hypothesized that social dominance orientation would be positively associated with online and offline hate speech perpetration (Hypothesis 1).

Social dominance orientation tends to be stable rather than changing over time and less malleable than other individual characteristics (Kleppesø et al., 2019). Thus, prevention and intervention efforts can benefit from exploring other variables that help to understand how hate speech relates to perpetration. Empathy and moral disengagement are associated with social dominance orientation (Sidanius et al., 2013) and hate speech perpetration (Wachs, Bilz, Wettstein, Wright, Kansok-Dusche, et al., 2022), and can be trained and increased by interventions in adolescence (Bustamante & Chaux, 2014) and adulthood (Soral et al., 2022). Hence, these two variables were examined with social dominance orientation and hate speech.

### 2.2 | Empathy, moral disengagement, and hate speech

Empathy refers to the ability to understand (cognitive component) and feel (affective component) another person's emotional state

(Batson, 2009). Moral disengagement is a socio-cognitive process that includes several mechanisms to separate oneself from actual or planned harmful behavior and avoid negative feelings such as guilt or shame (Bandura et al., 1996). According to Bandura, people do not usually perpetrate harm against others unless having justified to themselves “the morality of their actions” (1999, p. 194). Examples are blaming or dehumanizing victims (e.g., it is their fault they are rats) and diffusing responsibility (e.g., everyone does it; Bandura et al., 1996).

Previous research demonstrated that low levels of empathy and high levels of moral disengagement ease the perpetration of aggression and violence (Bandura, 1999; Zych et al., 2019). Specifically, while empathy buffers, moral disengagement strengthens the positive association between victimization and the perpetration of online hate speech (Wachs, Bilz, Wettstein, Wright, Kansok-Dusche, et al., 2022; Wachs, Bilz, Wettstein, Wright, Krause, et al., 2022). To our knowledge, previous studies did not simultaneously include online and offline contexts to address the complex associations between empathy, moral disengagement, and hate speech perpetration. Therefore, we aimed to fill this gap by these two settings in a single analysis.

### 2.3 | Empathy and moral disengagement as mediators of the association between social dominance orientation and hate speech perpetration

Bringing together the findings above, we argue that empathy and moral disengagement can be seen as bridges between social dominance orientation and hate speech perpetration. According to the socio-cognitive theory of moral agency (Bandura, 1986, 1999), people act by their moral reasoning, and moral reasoning translates into action (i.e., moral agency) using self-regulation (Bandura & Jourden, 1991). This process involves two self-regulatory mechanisms: the inhibitive, which prevents from acting immorally, and the proactive, which incentives to act morally (Bandura, 1999, 2002). Both mechanisms have an *anticipatory* nature, such that individuals activate these two self-regulatory mechanisms before behaving in a particular way (Bandura, 1999).

Against this backdrop, empathy can be considered an inhibitive factor in the association between social dominance orientation (the reasoning) and hate speech perpetration (the agency). Indeed, Bandura (1999) posed that individuals are likely to avoid harming others when empathy arises from recognizing shared humanity in others. Moreover, empathy and social dominance orientation are negatively related (Pratto et al., 1994; Sidanius et al., 2013; Sidanius & Pratto, 1999), and empathy mediates the association between social dominance orientation and prejudice during adulthood (Nicol & Rounding, 2013). In line with the socio-cognitive theory of moral agency and the evidence from adult samples, we hypothesized that social dominance orientation would have a statistically significant indirect effect on hate speech via low levels of empathy (Hypothesis 2).

In contrast to empathy, there is not much research considering the association between moral disengagement and social dominance orientation. Nonetheless, the socio-cognitive theory poses that moral disengagement serves the purpose of cognitively reinterpreting immoral reasoning into justifiable actions. As such, it acts as a proactive, anticipatory mechanism of self-regulation (Bandura, 2002). Therefore, social dominance orientation might result in hate speech perpetration if mechanisms of moral disengagement are activated. The bullying and aggression literature findings show that blaming the victims is positively associated with perpetration and pro-bullying support (Bjärehed et al., 2020). Moreover, dehumanizing and blaming victims mediate the association between social dominance orientation and attitudes supporting war (Jackson & Gaertner, 2010). In line with previous evidence and the socio-cognitive theory of moral agency, we hypothesized that social dominance orientation would have a significant indirect effect on hate speech via high levels of moral disengagement (Hypothesis 3).

## 2.4 | Gender differences

Although previous studies support that men score higher than women on social dominance orientation in adulthood (Sidanius et al., 2017; Zubielevitch et al., 2023), this gender difference is not present in childhood and adolescence (Pan et al., 2020; Volk et al., 2021). Furthermore, girls show higher levels of empathy, especially on its affective component (Farrell & Vaillancourt, 2021; Garandeau et al., 2022), lower levels of moral disengagement (Falla et al., 2021), and less frequency of online (Wachs et al., 2021) and offline hate speech perpetration (Castellanos et al., 2023; Kansok-Dusche et al., 2023). Regarding the association among these characteristics, studies show that boys exhibit stronger associations between social dominance orientation and physical aggression (Gumpel & Godtiner, 2023), moral disengagement (Gumpel & Godtiner, 2023), and bullying (Pan et al., 2020). In contrast, the negative association between empathy and the perpetration of bullying and cyberbullying seems not to vary by gender (Del Rey et al., 2016; Zych et al., 2019). In sum, the evidence about the gender variations of the potential associations examined in the present study inconclusive. Considering this, we analyzed if the direct and indirect effects tested in the present study varied by gender from an exploratory perspective. Therefore, we did not formulate *a priori* hypotheses concerning differences between boys and girls.

## 3 | METHOD

### 3.1 | Participants

Participants included 3225 adolescents from Germany ( $n = 1841$ ; 51.7%) and Switzerland ( $n = 1719$ ; 48.3%), in grades 7 ( $n = 1070$ ; 33.1%), 8 ( $n = 1147$ ; 35.6%), and 9 ( $n = 1008$ ; 31.3%), from 236

classes in 40 schools. Typically, students' age ranged between 13 and 15 years in these grades. In terms of gender, 46.1% ( $n = 1487$ ) self-identified as boys, 51.7% ( $n = 1668$ ) as girls, 2% ( $n = 64$ ) as gender diverse, and 0.2% ( $n = 6$ ) did not indicate their gender. Moreover, 37.2% ( $n = 1200$ ) of the participants had an immigrant background (i.e., they or one of their parents was born in a country other than Germany or Switzerland), and 62.8% ( $n = 2025$ ) had no immigrant background. From the participants, 30.8% ( $n = 994$ ) reported living in families of low affluence, 35.8% ( $n = 1155$ ) in families of medium affluence, 32.4% ( $n = 1046$ ) in families of high affluence, and 0.9% ( $n = 30$ ) had missing values on the socioeconomic status (SES) items.

### 3.2 | Procedure and sampling technique

Approval for this study was obtained from the data protection officer, the educational authority of the Federal State of Berlin and Brandenburg (Germany) and the University of Potsdam Ethics Committee (UP65/2018). All schools were initially stratified by federal state and type of school (e.g., grammar secondary school or nonacademic-track secondary school). After, for each school type, schools were randomly selected proportionally to their size to guarantee that all students had the same likelihood of being included in the sample regardless of the size of their school. In Germany, the acquisition pool of sample schools was composed of a stratified and randomized probability-proportional-to-size scheme (Yates & Grundy, 1953). In Switzerland, the acquisition pool of sample schools was designed via a contrastive sampling scheme based on rural/urban geography and high/low immigrant background.

From the resulting acquisition pools, a total of 100 schools (Germany:  $n = 76$ , Switzerland:  $n = 24$ ) were invited to participate in the study, and 40 of them agreed to take part (Germany:  $n = 18$ , Switzerland:  $n = 22$ ). Acquisition stopped as soon as the required sample size was attained. A total of 264 seventh, eighth, ninth, and mixed (Swiss classes in which students between 14 and 16 years old are taught together) classes were invited to participate (Germany:  $n = 106$ ; Switzerland  $n = 158$ ), and 89% of these ( $n = 236$ ) took part in the study (Germany:  $n = 98$ ; Switzerland:  $n = 138$ ). Of 5836 students, 3560 (Germany:  $n = 1841$ ; Switzerland:  $n = 1719$ ) participated in the study. Parents or legal guardians provided consent for all of them except those aged 14 and older in Berlin. For them, parents provided a written declaration saying they were informed about the study's aims, contents, and procedures and authorized their child to decide to participate and provide written consent voluntarily. This practice follows the legal regulation in the state, which poses that at this age, students have the legal autonomy to decide about participating in research studies. Participants from mixed classes ( $n = 335$ ) were not included in the present study since being in these classes could be confounded with being in a Swiss school. Students completed a survey on tablets (Swiss sample) or their own electronic devices (German sample).

### 3.3 | Measures

#### 3.3.1 | Hate speech perpetration

At first, participants were presented with a definition of hate speech as a short video clip, presented on their own electronic device, followed by a brief text-based introduction emphasizing that hate speech is intentionally hurtful, occurs in public, is directed toward social groups, and can take nonverbal forms (see Supporting Information Material). After, students were asked to report how often they perpetrated offline (i.e., "hate speech happening in school without the use of digital media") and online hate speech (i.e., "hate speech on the Internet") in the previous 12 months, by using a five-point scale that ranged from "never," to "several times a week."

#### 3.3.2 | Social dominance orientation

An instrument from Klocke (2012) was adapted by adding an introductory text asking students to think about different social groups with specific examples (see details on Supporting Information Material). Then, students rated their level of agreement with eight statements (e.g., "It is probably ok that certain groups are at the top of society and others at the bottom"), using a five-point scale that ranged from "absolutely disagree" (1) to "absolutely agree" (5). Cronbach's  $\alpha$  was .79, and McDonald's  $\omega$  was .78.

#### 3.3.3 | Empathy for victims of hate speech and moral disengagement

An instrument from Knauf et al. (2018) was adapted by referring to hate speech in the introduction: "When I see classmates being insulted or attacked by other classmates because of their skin color, origin, religion, sexual orientation, or gender...". Six items assessed different aspects of empathy (e.g., "I realize how badly they are doing"), and nine items assessed moral disengagement mechanisms (e.g., attribution of blame: "I figure it is their fault"). Items were rated on a five-point scale from "absolutely disagree" (1) to "absolutely agree" (5). Cronbach's  $\alpha$  for empathy was .91 and .89 for moral disengagement. McDonald's  $\omega$  was .91 for empathy and .89 for moral disengagement.

#### 3.3.4 | Sociodemographic variables

Participants were asked for their grade and gender (boy, girl, gender diverse). The immigrant background was assessed by asking whether the participants or one of their parents was born in a country other than Germany or Switzerland. SES was measured using the Family Affluence Scale (Hartley et al., 2016), which includes questions about family possessions (e.g., family's car).

### 3.4 | Data analyses

#### 3.4.1 | Power analysis and missing data

A priori conducted power analysis with G\*Power (Faul et al., 2007) set to  $\alpha = .05$  and power = .80 revealed that the required sample size was at least 782 participants to detect small to medium correlational effect sizes. Anticipating nonresponse and accounting for the nested structure of the sample, the required minimum sample size was  $N = 1944$  students in 108 classes at 18 schools (Teerenstra et al., 2010). Accordingly, the present sample size was adequate to investigate our hypotheses. Overall, missing data were between 1% and 1.9%. Little's MCAR test revealed that the data were not missing completely at random ( $\chi^2 [26] = 41.94$ ,  $p = .025$ ). Consequently, missing data were handled using the full information maximum likelihood approach (Muthén & Muthén, 2012–2021).

#### 3.4.2 | Statistical analyses

To test the direct and indirect effects of social dominance orientation on hate speech perpetration, a two-level mediation path model was built in MPlus 8.7 (Muthén & Muthén, 2012–2021), with students at level 1 (L1) nested within classrooms at level 2 (L2). The model included direct and indirect effects to test mediations only at L1 (i.e., 1-1-1 model; Preacher et al., 2010).

Separate analyses were conducted for empathy and moral disengagement in three steps. First, a model with a random intercept only to estimate the intraclass correlation of the outcome variables of the model: offline hate speech perpetration, online hate speech perpetration, and empathy (Model 0). The subsequent paths were included as fixed effects, as our main interest was the associations at L1 (Snijders & Bosker, 2016). In Model 1, the control variables grade, immigrant background, and SES at L1 were included as predictors of the outcomes—dummy-coded variables with an immigrant background and ninth grade as reference categories were included. In Model 2, direct and indirect effects were tested simultaneously. To estimate direct effects, we included social dominance orientation and empathy as predictors of offline and online hate speech perpetration and a path of empathy regressed on social dominance orientation. The indirect effects of social dominance orientation on offline and online hate speech perpetration via empathy were computed by the command "INDIRECT EFFECTS." The same three steps were replicated with models that included moral disengagement instead of empathy. Finally, we conducted a moderated mediation analysis to explore if gender moderated the direct and indirect effects tested previously. For this analysis, the group of students who self-identified as gender diverse was not included, given the small sample size.

Direct and indirect effects were assessed using a statistical significance test. Given the multilevel nature of the data, model improvement was assessed by relative decreases in Akaike's

information criterion (AIC; Akaike, 1974). At least 10 points of relative reductions on the AIC is considered an improvement of a model with respect to a previous one (Burnham & Anderson, 2004). Finally, relative increases in the explained variance ( $R^2$ ) of offline and online hate speech perpetration were also considered to indicate a better model fit.

## 4 | RESULTS

### 4.1 | Descriptive statistics and intraclass correlation coefficients

Descriptive statistics and bivariate correlations between the study variables are displayed in the Supporting Information Material. Confirming our first hypothesis, social dominance orientation was positively associated with online ( $r = .26, p < .001$ ) and offline ( $r = .26, p < .001$ ) hate speech perpetration. The intraclass correlation coefficients were .10 for offline hate speech perpetration and .05 for online hate speech perpetration. Therefore, it was adequate to proceed with the multilevel analysis.

### 4.2 | Direct and indirect effects of social dominance orientation on hate speech perpetration via empathy

Model 0 provided the initial values for the subsequent relative model fit comparisons (AIC = 23,588.84). Table 1 displays the standardized coefficients of the multilevel model on the prediction of online and offline hate speech perpetration. The analyses revealed that SES, immigrant background, and grade were not statistically significant predictors of offline and online hate speech perpetration. With the addition of these control variables, a reduction of 1091.27 absolute units on the AIC (4.63%) was observed, suggesting a better fit of this model with respect to Model 0. Findings from Model 2 provided evidence of the direct effects of social dominance orientation on offline ( $\beta = .22, 95\% \text{ CI: } [0.17-0.26], p < .001$ ) and online ( $\beta = .22, 95\% \text{ CI: } [0.18-0.26], p < .001$ ) hate speech perpetration, confirming our first hypothesis. Furthermore, three direct effects were statistically significant. First, the direct effect of social dominance orientation on empathy ( $\beta = -.46, 95\% \text{ CI: } [-0.50 \text{ to } -0.44], p < .001$ ), the direct effect of empathy on offline hate speech perpetration ( $\beta = -.10, 95\% \text{ CI: } [-0.14 \text{ to } -0.06], p < .001$ ), and the direct effect of empathy on

**TABLE 1** Direct and indirect effects of social dominance orientation on hate speech perpetration via empathy.

Predictor	Model 1 (control variables)			Model 2 (direct and indirect effects)		
	$\beta$ [95% CI]	SE	$p$	$\beta$ [95% CI]	SE	$p$
Outcome: Offline hate speech perpetration						
SES	.02 [-0.01 to 0.05]	0.02	.290	-.01 [-0.04 to 0.03]	0.02	.790
Immigrant background	.03 [-0.01 to 0.06]	0.02	.217	.01 [-0.02 to 0.05]	0.02	.465
7th grade	-.01 [-0.06 to 0.05]	0.03	.874	.00 [-0.05 to 0.05]	0.03	.996
8th grade	.01 [-0.05 to 0.07]	0.04	.771	.02 [-0.03 to 0.07]	0.03	.522
Empathy				-.10 [-0.14 to -0.06]	0.02	<.001
Social dominance orientation (direct)				.22 [0.17 to 0.26]	0.03	<.001
Social dominance orientation (indirect)				.05 [0.03 to 0.07]	0.01	<.001
Outcome: Online hate speech perpetration						
SES	.02 [-0.02 to 0.06]	0.03	.429	.00 [-0.04 to 0.03]	0.02	.842
Immigrant background	.02 [-0.01 to 0.05]	0.02	.318	.02 [-0.01 to 0.05]	0.02	.376
7th grade	-.04 [-0.09 to 0.01]	0.03	.154	-.03 [-0.06 to 0.01]	0.02	.277
8th grade	-.03 [-0.07 to 0.02]	0.03	.377	-.01 [-0.05 to 0.03]	0.02	.801
Empathy				-.09 [-0.13 to -0.05]	0.02	<.001
Social dominance orientation (direct)				.22 [0.18 to 0.26]	0.02	<.001
Social dominance orientation (indirect)				.04 [0.02 to 0.06]	0.02	<.001
AIC	22,497.57			21,406.60		
$R^2$ Offline	0.01%			7.7%		
$R^2$ Online	0.02%			7.7%		

Note: L1 = Reference categories for immigrant background: no immigrant background; for grade: ninth grade.

Abbreviations: AIC, Akaike's information criterion; SES, socioeconomic status.

online hate speech perpetration ( $\beta = -.09$ , 95% CI:  $[-0.13$  to  $-0.05]$ ,  $p \leq .001$ ). Finally, the indirect effects of social dominance orientation on offline ( $\beta = .05$ , 95% CI:  $[0.03$ – $0.07]$ ,  $p \leq .001$ ) and online hate speech perpetration ( $\beta = .04$ , 95% CI:  $[0.02$ – $0.06]$ ,  $p \leq .001$ ), via empathy, were statistically significant, confirming our second hypothesis. Adding these direct and indirect effects resulted in a better fit of the model, as evidenced by a decrease of 1090.97 units (4.85%) on the AIC with respect to Model 1.

### 4.3 | Direct and indirect effects of social dominance orientation on hate speech perpetration, via moral disengagement

Model 0 provided the initial values for the subsequent relative model fit comparisons (AIC = 22,136.52). Table 2 displays the standardized coefficients of the multilevel model on the prediction of online and offline hate speech perpetration. Similar to the findings from the models of empathy, Model 1 showed that SES, immigration background, and grade had no association with hate speech perpetration (Model 1). With the addition of these control variables, the AIC was reduced by 1092.97 units (4.94%), suggesting a better fit of this model with respect to Model 0. Findings from Model 2 provided

evidence of the direct effects of social dominance orientation on offline ( $\beta = .17$ , 95% CI:  $[0.13$ – $0.22]$ ,  $p < .001$ ) and online ( $\beta = .17$ , 95% CI:  $[0.13$ – $0.22]$ ,  $p < .001$ ) hate speech perpetration. Furthermore, three direct effects were statistically significant. First, the direct effect of social dominance orientation on moral disengagement ( $\beta = .53$ , 95% CI:  $[0.50$ – $0.56]$ ,  $p < .001$ ), the direct effect of moral disengagement on offline hate speech perpetration ( $\beta = .17$ , 95% CI:  $[0.13$ – $0.22]$ ,  $p < .001$ ), and the direct effect of moral disengagement on online hate speech perpetration ( $\beta = .17$ , 95% CI:  $[0.13$ – $0.22]$ ,  $p = .004$ ). Finally, the indirect effects of social dominance orientation on offline ( $\beta = .09$ , 95% CI:  $[0.07$ – $0.11]$ ,  $p < .001$ ) and online hate speech perpetration ( $\beta = .09$ , 95% CI:  $[0.06$ – $0.11]$ ,  $p < .001$ ), via moral disengagement were statistically significant, confirming our third hypothesis. The addition of the direct and indirect effects tested in Model 2 resulted in a better fit than Model 1, as evidenced by a decrease of 1382.37 units (6.57%) in the AIC.

### 4.4 | Moderated mediation

The model parameters of the moderated mediation analyses are shown in Table 3. As displayed, the differences between boys and girls on the direct, indirect, and total effects were not statistically

**TABLE 2** Direct and indirect effects of social dominance orientation on hate speech perpetration via moral disengagement.

Predictor	Model 1 (control variables)			Model 2 (direct and indirect effects)		
	$\beta$ [95% CI]	SE	$p$	$\beta$ [95% CI]	SE	$p$
Outcome: Offline hate speech perpetration						
SES	.02 $[-0.01$ to $0.05]$	0.02	.290	-.01 $[-0.03$ to $0.03]$	0.02	.799
Immigrant background	.03 $[-0.01$ to $0.06]$	0.02	.217	.00 $[-0.03$ to $0.03]$	0.02	.829
7th grade	-.01 $[-0.06$ to $0.05]$	0.03	.874	.00 $[-0.04$ to $0.05]$	0.03	.905
8th grade	.01 $[-0.05$ to $0.07]$	0.04	.771	.02 $[-0.03$ to $0.07]$	0.03	.428
Moral disengagement				.17 $[0.13$ to $0.21]$	0.02	<.001
Social dominance orientation (direct)				.17 $[0.13$ to $0.22]$	0.03	<.001
Social dominance orientation (indirect)				.09 $[0.07$ to $0.11]$	0.01	<.001
Outcome: Online hate speech perpetration						
SES	.02 $[-0.02$ to $0.06]$	0.03	.429	.00 $[-0.04$ to $0.03]$	0.02	.846
Immigrant background	.02 $[-0.01$ to $0.05]$	0.02	.318	.01 $[-0.02$ to $0.04]$	0.02	.728
7th grade	-.04 $[-0.09$ to $0.01]$	0.03	.154	-.02 $[-0.06$ to $0.02]$	0.02	.344
8th grade	-.03 $[-0.07$ to $0.02]$	0.03	.377	.00 $[-0.04$ to $0.04]$	0.02	.940
Moral disengagement				.17 $[0.13$ to $0.21]$	0.03	<.001
Social dominance orientation (direct)				.17 $[0.13$ to $0.22]$	0.03	<.001
Social dominance orientation (indirect)				.09 $[0.06$ to $0.11]$	0.01	<.001
AIC	21,043.55			19,661.18		
R <sup>2</sup> Offline	0.01%			9%		
R <sup>2</sup> Online	0.02%			9.1%		

Note: L1 = Reference categories for immigrant background: no immigrant background; for grade: ninth grade.

Abbreviations: AIC, Akaike's information criterion; SES, socioeconomic status.



**TABLE 3** Gender differences on the direct and indirect effects of social dominance orientation on hate speech perpetration. Results of moderated mediation.

Mediator/group parameter effect	Offline hate speech perpetration			Online hate speech perpetration		
	Estimate [95% CI]	SE	p	Estimate [95% CI]	SE	p
Empathy as mediator						
Boys						
Indirect	0.02 [−0.01 to 0.05]	0.02	.202	0.04 [0.02 to 0.07]	0.01	.001
Direct	0.29 [0.20 to 0.38]	0.06	<.001	0.23 [0.16 to 0.30]	0.04	<.001
Total effects	0.31 [0.23 to 0.40]	0.05	<.001	0.27 [0.20 to 0.34]	0.04	<.001
Girl						
Indirect	0.06 [0.04 to 0.09]	0.02	<.001	0.02 [0.00 to 0.04]	0.01	.068
Direct	0.20 [0.14 to 0.26]	0.04	<.001	0.18 [0.13 to 0.24]	0.04	<.001
Total effects	0.27 [0.20 to 0.33]	0.04	<.001	0.20 [0.15 to 0.26]	0.03	<.001
Difference in effects (boys–girls)						
Indirect	−0.04 [−0.08 to −0.01]	0.02	.062	0.03 [0.00 to 0.05]	0.02	.141
Direct	0.09 [−0.01 to 0.19]	0.06	.156	0.04 [−0.05 to 0.13]	0.06	.456
Total effects	0.05 [−0.05 to 0.15]	0.06	.425	0.07 [−0.02 to 0.16]	0.06	.227
Moral disengagement as mediator						
Boys						
Indirect	0.008 [0.04 to 0.12]	0.02	<.001	0.07 [0.04 to 0.10]	0.02	<.001
Direct	0.23 [0.14 to 0.31]	0.05	<.001	0.19 [0.12 to 0.27]	0.05	<.001
Total effects	0.31 [0.22 to 0.39]	0.05	<.001	0.27 [0.19 to 0.34]	0.04	<.001
Girls						
Indirect	0.11 [0.07 to 0.15]	0.02	<.001	0.08 [0.04 to 0.11]	0.02	<.001
Direct	0.16 [0.10 to 0.22]	0.04	<.001	0.13 [0.08 to 0.17]	0.03	<.001
Total effects	0.27 [0.20 to 0.33]	0.04	<.001	0.21 [0.15 to 0.26]	0.03	<.001
Difference in effects (boys–girls)						
Indirect	−0.03 [−0.08 to 0.02]	0.03	.357	−0.01 [−0.05 to 0.04]	0.03	.837
Direct	0.07 [−0.03 to 0.17]	0.06	.259	0.07 [−0.02 to 0.15]	0.05	.227
Total effects	0.04 [−0.06 to 0.14]	0.06	.502	0.06 [−0.03 to 0.15]	0.06	.277

significant. These findings indicate that social dominance orientation, directly and indirectly, affected hate speech perpetration for both boys and girls to the same extent.

## 5 | DISCUSSION

In the present study, we aimed to investigate the association between social dominance orientation and adolescents' online and offline hate speech perpetration and whether social dominance orientation indirectly affected hate speech perpetration via empathy and moral disengagement. In line with the social dominance theory (Pratto et al., 1994; Sidanius & Pratto, 1999), our first hypothesis posited that social dominance orientation would be

positively associated with online and offline hate speech perpetration. The findings confirmed this prediction. Although our instrument was adapted to measure social dominance orientation in general, rather than to differentiate among the stratifications systems proposed by the social dominance theory, a preference of hate speech perpetrators for the arbitrary system could explain this direct effect. This system is based on groups defined arbitrarily, under the biased impression of social distinctions related to power (e.g., for ethnicity, religion, or nationality, among others; Pratto et al., 2006). Accordingly, the perpetration of hate speech would be eased for students who consider their in-group better or superior to others. Previous studies showed similar results for adult samples (Amiot & Bourhis, 2005; Bilewicz et al., 2017; Kteily et al., 2012). This study contributes to the literature by providing evidence for a

sample of adolescents for the case of offline and online hate speech perpetration.

### 5.1 | Inhibitive and proactive mechanisms of self-regulation in social dominance orientation and hate speech perpetration

Our second and third hypotheses posited that two indirect effects of social dominance orientation on online and offline hate speech perpetration via empathy and moral disengagement would be observed. The evidence supported these predictions. These findings exemplify the dual nature of self-regulation, as proposed in Bandura's theory (1986, 1999). Although empathy acted as an inhibitory mechanism, moral disengagement acted as a proactive mechanism that facilitated the perpetration of hate speech by adolescents with a strong orientation for maintaining group hierarchies. Our results mirror findings from adult samples in related fields such as prejudice (Nicol & Rounding, 2013) and attitudes favoring war (Jackson & Gaertner, 2010). However, our effects were small. A possible explanation is the age of our participants. According to Bandura (1999), moral disengagement is "gradualistic." Individuals perform "milder aggressive acts that they can tolerate with some discomfort," but immoral acts can, with repetition, be routinized (Bandura, 1999, p. 203). In our sample, most students did not perpetrate hate speech in the previous 12 months (i.e., 78.7% for offline and 87.2% for online) and did show low levels of social dominance orientation. It is possible that with age, social dominance orientation is reinforced. A study with early adolescents found that participants who increased their moral disengagement over 4 years showed steeper growth trajectories of verbal bullying perpetration (Bjärehed et al., 2020). Moreover, a longitudinal study that followed participants from adolescence (16–18 years) to young adulthood (25 years) showed that although moral disengagement predicted aggressive behavior over time, this association became stronger with age (Caprara et al., 2014). Therefore, only by then inhibitive and proactive self-regulation mechanisms would be required and activated more easily. This proposition requires empirical testing using longitudinal studies with follow-up measurements from adolescence to adulthood.

Our study is novel regarding the simultaneous inclusion of online and offline forms of hate speech perpetration. Although some overlap was expected, the analyses revealed a high similarity between these two settings. Social dominance, empathy, and moral disengagement predicted online and offline forms of hate speech perpetration similarly. Moreover, we found that offline and online hate speech perpetration were moderately correlated. Therefore, it is plausible to think that the correlates of hate speech perpetration in one of these settings might ease the perpetration in the other, as it has been observed in other forms of aggressive behavior during adolescence, such as bullying and cyberbullying (Kim et al., 2022) and offline and online dating violence (Cava et al., 2020). Furthermore, more specific components of empathy and moral disengagement might evidence some differences. From related research fields, there

is evidence of differential associations between the components of affective empathy and bullying perpetration. For instance, peripheral responsivity (i.e., a vicarious, as opposed to direct, experience of feeling another's emotions) has a negative association with cyberbullying but not with offline bullying (Graf et al., 2019).

Acknowledging the previously documented gender differences in the constructs examined in our study, we examined if the direct and indirect effects observed varied by gender. The evidence showed that the results were the same for boys and girls. Evidence from adult samples might help understand this finding. Although men tend to score higher on measures of social dominance than women (Sidanius et al., 2017), some studies show that this gender difference is better explained by conformity with traditional and stereotypical gender roles rather than gender itself (Aranda et al., 2015; Wilson & Liu, 2003). Thus, the effect of social dominance orientation on hate speech perpetration might differ depending on the extent to which adolescents adhere to social expectations of traditional gender roles. This explanation requires further empirical exploration. Our findings add to the current debate about the gender differences in the associations among the variables studied in this study. Our findings align with those on empathy and aggressiveness (Del Rey et al., 2016; Zych et al., 2019), but differ from those on moral disengagement, social dominance, and aggression (Gumpel & Gotdiner, 2023; Pan et al., 2020).

### 5.2 | Applications for prevention of hate speech from schools

Anti-hate speech interventions might benefit from our findings. Although social dominance orientation seems less malleable to ambient (as it is more of a stable personal characteristic; Klepeš et al., 2019), empathy and moral disengagement can be modified with educational strategies. For instance, the empirically tested school program *HateLess* increased empathy toward victims of hate speech by fostering intergroup contact (e.g., indirect contact using movies or stories), knowledge (e.g., information about negative consequences for victims' well-being), and development of skills (e.g., role-plays to understand and empathize with targeted minorities) (Wachs et al., 2023). Furthermore, prevention science evidence shows that interventions to develop critical thinking might reduce moral disengagement (Bustamante & Chaux, 2014).

Accordingly, one can expect that combined strategies targeted to decrease moral disengagement and foster empathy would not only directly impact hate speech perpetration but also would decrease the probability of individuals with a high social dominance orientation perpetrating it. Humanization is a powerful process that fosters empathetic responses and decreases moral disengagement simultaneously, which can be understood as "the affirmation of common humanity" (Bandura, 2002, p. 110).

These strategies must be adapted to the context of hate speech. First, individuals are more likely to be empathetic with members of their in-group than with members of an out-group (Fuchs, 2019).



In this regard, prevention programs can focus on fostering empathy in situations of intergroup conflict and extending empathetic responses to members other than their own social group. Second, some mechanisms of moral disengagement have a stronger component of social interactions than others. For instance, the diffusion of responsibility occurs, by definition, in a collective, where perpetrators can attribute blame and responsibility to members of their group. Thus, prevention programs could focus on providing students with strategies to deal with moral disengagement mechanisms related to group dynamics.

Finally, in line with Bandura's triadic codetermination theory (2018), which poses that human agency is the product of personal, behavioral, and environmental determinants, schools can promote strategies targeted at multiple levels of influence. The personal level (e.g., promoting empathy), the interpersonal level (e.g., promoting countering hate speech), and the contextual level (e.g., promoting prosocial school climate). Indeed, empirically tested school programs documented the positive impact of combining multiple levels of intervention for reducing hate speech (e.g., *HateLess*; Wachs et al., 2023), bullying (KiVa; Garandeau et al., 2022), and cyberbullying (*Medienhelden* [Media Heroes]; Schultze-Krumbholz et al., 2018).

### 5.3 | Limitations and future research

Despite our novel examination of the direct and indirect effects of social dominance orientation on online and offline hate speech perpetration, the present study is not exempt from limitations. First, we did not include any contextual variables in our analysis. Bandura's triadic codetermination theory (2018) poses that human agency is the product of personal, behavioral, and environmental determinants. In the present study, we controlled for the nested nature of classrooms, but we did not include any other information regarding the classroom. Future studies should advance in this direction.

Another limitation is related to directionality. In line with the socio-cognitive theory, the evidence supported our claim that social dominance orientation, empathy, and moral disengagement precede offline and online hate speech perpetration. Longitudinal studies have demonstrated this direction in adults (Amiot & Bourhis, 2005; Kteily et al., 2012). However, our findings are based on a cross-sectional design limiting the possibility of interpreting the results as causal. Hate speech perpetration might also impact empathy and moral disengagement and reinforce social dominance orientation. For instance, while dehumanization eases exerting aggression, diffusion of responsibility, or attribution of the blame can emerge as the result of a reasoning process after an immoral behavior is committed. Retrospective studies with adults demonstrated that justifications of this type came after committing highly violent acts (Bandura, 2002). For this, measures of specific components of social dominance orientation, empathy, and moral disengagement are required, as well as longitudinal designs. Finally, using a single item to assess perpetration and self-report questionnaires were limitations. Future studies could employ other methodologies to reduce social desirability, such as peer nominations (Bukowski et al., 2017).

## 6 | CONCLUSION

Although social dominance orientation positively correlates with hate speech perpetration, not all individuals with an orientation for social hierarchies engage in devaluating a social group. To better understand this, we examined whether social dominance orientation also indirectly affected hate speech perpetration via two self-regulatory mechanisms, empathy, and moral disengagement. The findings confirmed our hypotheses. First, we observed a positive association between social dominance orientation and online and offline hate speech perpetration. Second, social dominance orientation was associated with hate speech perpetration via low levels of empathy and high levels of moral disengagement. Although small, the effects constitute important evidence to add to the literature on hate speech in adolescence and to elucidate paths of action for preventing hate speech in the school context. Hate speech and bullying are different phenomena (Kansok-Dusche et al., 2022), but similar moral and emotional correlates explain the perpetration of these behavior. Therefore, our conclusions and recommendations can be applied to preventing aggression in general.

### ACKNOWLEDGMENTS

This work was supported by the University of Teacher Education Bern [grant number 19s 0008 01], and the German Research Foundation (DFG) [grant numbers WA 4275/2-1 and BI 1046/9-1]. The research funding sources had no involvement in the conduct of the project or preparation of the present manuscript. Open access funding provided by Pädagogische Hochschule Bern.

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

### DATA AVAILABILITY STATEMENT

Research data are not shared.

### REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Amiot, C. E., & Bourhis, R. Y. (2005). Ideological beliefs as determinants of discrimination in positive and negative outcome distributions. *European Journal of Social Psychology*, 35(5), 581–598. <https://doi.org/10.1002/ejsp.238>
- Aranda, M., Montes-Berger, B., & Castillo-Mayén, M.-R. (2015). The social dominance orientation in adolescence: the role of gender identity and stereotypical male and female traits [La orientación a la dominancia social en la adolescencia: el papel de la identidad de género y los rasgos estereotípicos femeninos y masculinos]. *Revista de Psicología Social*, 30(2), 241–263. <https://doi.org/10.1080/21711976.2015.1016752>
- Bäckström, M., & Björklund, F. (2007). Structural modeling of generalized prejudice: The role of social dominance, authoritarianism, and empathy. *Journal of Individual Differences*, 28(1), 10–17. <https://doi.org/10.1027/1614-0001.28.1.10>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193–209. [https://doi.org/10.1207/s15327957pspr0303\\_3](https://doi.org/10.1207/s15327957pspr0303_3)

- Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, 31(2), 101–119. <https://doi.org/10.1080/0305724022014322>
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, 71(2), 364–374. <https://doi.org/10.1037/0022-3514.71.2.364>
- Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology*, 60(6), 941–951. <https://doi.org/10.1037/0022-3514.60.6.941>
- Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–16). The MIT Press. <https://doi.org/10.7551/mitpress/9780262012973.003.0002>
- Bilewicz, M., Soral, W., Marchlewska, M., & Winiewski, M. (2017). When authoritarians confront prejudice. Differential effects of SDO and RWA on support for hate-speech prohibition: When authoritarians confront prejudice. *Political Psychology*, 38(1), 87–99. <https://doi.org/10.1111/pops.12313>
- Bjårehed, M., Thornberg, R., Wänström, L., & Gini, G. (2020). Mechanisms of moral disengagement and their associations with indirect bullying, direct bullying, and pro-aggressive bystander behavior. *The Journal of Early Adolescence*, 40(1), 28–55. <https://doi.org/10.1177/0272431618824745>
- Bukowski, W. M., Castellanos, M., & Persram, R. J. (2017). The current status of peer assessment techniques and sociometric methods: The current status of peer assessment techniques and sociometric methods. *New Directions for Child and Adolescent Development*, 2017(157), 75–82. <https://doi.org/10.1002/cad.20209>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Bustamante, A., & Chau, E. (2014). Reducing moral disengagement mechanisms: A comparison of two interventions. *Journal of Latino/Latin American Studies*, 46(1), 52–54. <https://doi.org/10.18085/llas.6.1.123583644qq115t3>
- Caprara, G. V., Tisak, M. S., Alessandri, G., Fontaine, R. G., Fida, R., & Paciello, M. (2014). The contribution of moral disengagement in mediating individual tendencies toward aggression and violence. *Developmental Psychology*, 50(1), 71–85. <https://doi.org/10.1037/a0034488>
- Castellanos, M., Wettstein, A., Wachs, S., Kansok-Dusche, J., Ballaschk, C., Krause, N., & Bilz, L. (2023). Hate speech in adolescents: A binational study on prevalence and demographic differences. *Frontiers in Education*, 8, 1–14. <https://doi.org/10.3389/educ.2023.1076249>
- Cava, M.-J., Martínez-Ferrer, B., Buelga, S., & Carrascosa, L. (2020). Sexist attitudes, romantic myths, and offline dating violence as predictors of cyber dating violence perpetration in adolescents. *Computers in Human Behavior*, 111, 106449. <https://doi.org/10.1016/j.chb.2020.106449>
- Falla, D., Romera, E. M., & Ortega-Ruiz, R. (2021). Aggression, moral disengagement and empathy: A longitudinal study within the interpersonal dynamics of bullying. *Frontiers in Psychology*, 12, 703468. <https://doi.org/10.3389/fpsyg.2021.703468>
- Farrell, A. H., & Vaillancourt, T. (2021). Adolescent empathic concern and perspective taking: Heterogeneous developmental trajectories and childhood social and psychological factors. *Journal of Personality*, 89(4), 672–688. <https://doi.org/10.1111/jopy.12607>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fuchs, T. (2019). Empathy, group identity, and the mechanisms of exclusion: An investigation into the limits of empathy. *Topoi*, 38(1), 239–250. <https://doi.org/10.1007/s11245-017-9499-z>
- Garandeau, C. F., Laninga-Wijnen, L., & Salmivalli, C. (2022). Effects of the KiVa anti-bullying program on affective and cognitive empathy in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 51(4), 515–529. <https://doi.org/10.1080/15374416.2020.1846541>
- Goodboy, A. K., Martin, M. M., & Rittenour, C. E. (2016). Bullying as a display of social dominance orientation. *Communication Research Reports*, 33(2), 159–165. <https://doi.org/10.1080/08824096.2016.1154838>
- Graf, D., Yanagida, T., & Spiel, C. (2019). Through the magnifying glass: Empathy's differential role in preventing and promoting traditional and cyberbullying. *Computers in Human Behavior*, 96, 186–195. <https://doi.org/10.1016/j.chb.2019.02.007>
- Gumpel, T. P., & Goddiner, V. (2023). The use of latent profile analyses to examine social dominance orientation as a predictor of peer-to-peer aggression among adolescents. *Journal of Interpersonal Violence*, 38(1–2), 1424–1445. <https://doi.org/10.1177/08862605221090572>
- Hartley, J. E. K., Levin, K., & Currie, C. (2016). A new version of the HBSC family affluence scale—FAS III: Scottish qualitative findings from the international FAS development study. *Child Indicators Research*, 9(1), 233–245. <https://doi.org/10.1007/s12187-015-9325-3>
- Jackson, L. E., & Gaertner, L. (2010). Mechanisms of moral disengagement and their differential use by right-wing authoritarianism and social dominance orientation in support of war. *Aggressive Behavior*, 36(4), 238–250. <https://doi.org/10.1002/ab.20344>
- Kansok-Dusche, J., Ballaschk, C., Krause, N., Zeißig, A., Seemann-Herz, L., Wachs, S., & Bilz, L. (2022). A systematic review on hate speech among children and adolescents: Definitions, prevalence, and overlap with related phenomena. *Trauma, Violence, & Abuse*, 1–18. <https://doi.org/10.1177/15248380221108070>
- Kansok-Dusche, J., Wettstein, A., Castellanos, M., Schwab, C., Subramaniam, A., Wachs, S., & Bilz, L. (2023). Associations between social competence, perceived parents' prosocial educational goals and adolescents' hate speech perpetration in school. *Victims & Offenders*, 1–18. <https://doi.org/10.1080/15564886.2023.2189191>
- Kim, J., Lee, Y., & Jennings, W. G. (2022). A path from traditional bullying to cyberbullying in South Korea: Examining the roles of self-control and deviant peer association in the different forms of bullying. *Journal of Interpersonal Violence*, 37(9–10), 5937–5957. <https://doi.org/10.1177/08862605211067022>
- Kleppeste, T. H., Czajkowski, N. O., Vassend, O., Røysamb, E., Eftedal, N. H., Sheehy-Skeffington, J., Kunst, J. R., & Thomsen, L. (2019). Correlations between social dominance orientation and political attitudes reflect common genetic underpinnings. *Proceedings of the National Academy of Sciences*, 116(36), 17741–17746. <https://doi.org/10.1073/pnas.1818711116>
- Klocke, U. (2012). Akzeptanz sexueller Vielfalt an Berliner Schulen: Eine Befragung zu Verhalten, Einstellungen und Wissen zu LSBT und deren Einflussvariablen. Humboldt Universität zu Berlin. [https://www.psychologie.huberlin.de/de/prof/org/download/klocke2012\\_1](https://www.psychologie.huberlin.de/de/prof/org/download/klocke2012_1)
- Knauf, R.-K., Eschenbeck, H., & Hock, M. (2018). Bystanders of bullying: Social-cognitive and affective reactions to school bullying and cyberbullying. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 12(4), article 3. <https://doi.org/10.5817/CP2018-4-3>
- Krause, N., Ballaschk, C., Schulze-Reichelt, F., Kansok-Dusche, J., Wachs, S., Schubarth, W., & Bilz, L. (2021). Ich lass mich da nicht klein machen! Eine qualitative Studie zur Bewältigung von Hate-speech durch Schüler/innen. *Zeitschrift für Bildungsforschung*, 11(1), 169–185. <https://doi.org/10.1007/s35834-021-00291-w>
- Kteily, N., Ho, A. K., & Sidanius, J. (2012). Hierarchy in the mind: The predictive power of social dominance orientation across social contexts and domains. *Journal of Experimental Social Psychology*, 48(2), 543–549. <https://doi.org/10.1016/j.jesp.2011.11.007>
- McFarland, S. (2010). Authoritarianism, social dominance, and other roots of generalized prejudice: psychological roots of generalized

- prejudice. *Political Psychology*, 31(3), 453–477. <https://doi.org/10.1111/j.1467-9221.2010.00765.x>
- Muthén, B., & Muthén, L. (2012–2021). *MPlus* (Version 8.7.) [Computer Software].
- Nicol, A. A. M., & Rounding, K. (2013). Alienation and empathy as mediators of the relation between social dominance orientation, right-wing authoritarianism and expressions of racism and sexism. *Personality and Individual Differences*, 55(3), 294–299. <https://doi.org/10.1016/j.paid.2013.03.009>
- Pan, B., Zhang, L., Ji, L., Garandeau, C. F., Salmivalli, C., & Zhang, W. (2020). Classroom status hierarchy moderates the association between social dominance goals and bullying behavior in middle childhood and early adolescence. *Journal of Youth and Adolescence*, 49(11), 2285–2297. <https://doi.org/10.1007/s10964-020-01285-z>
- Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology*, 17(1), 271–320. <https://doi.org/10.1080/10463280601055772>
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67(4), 741–763. <https://doi.org/10.1037/0022-3514.67.4.741>
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209–233. <https://doi.org/10.1037/a0020141>
- Del Rey, R., Lazuras, L., Casas, J. A., Barkoukis, V., Ortega-Ruiz, R., & Tsorbatzoudis, H. (2016). Does empathy predict (cyber) bullying perpetration, and how do age, gender and nationality affect this relationship. *Learning and Individual Differences*, 45, 275–281. <https://doi.org/10.1016/j.lindif.2015.11.021>
- Schultze-Krumbholz, A., Zagorscak, P., & Scheithauer, H. (2018). A school-based cyberbullying preventive intervention approach: The media heroes program. In M. Campbell & S. Bauman (Eds.), *Reducing cyberbullying in schools* (pp. 145–158). Elsevier. <https://doi.org/10.1016/B978-0-12-811423-0.00011-0>
- Sidanius, J., Cotterill, S., Sheehy-Skeffington, J., Kteily, N., & Carvacho, H. (2017). Social dominance theory: Explorations in the psychology of oppression. In C. G. Sibley & F. K. Barlow (Eds.), *The Cambridge handbook of the psychology of prejudice* (pp. 149–187). Cambridge University Press. <https://doi.org/10.1017/9781316161579.008>
- Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Ho, A. K., Sibley, C., & Duriez, B. (2013). You're inferior and not worth our concern: The interface between empathy and social dominance orientation: Empathy and social dominance orientation. *Journal of Personality*, 81(3), 313–323. <https://doi.org/10.1111/jopy.12008>
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139175043>
- Snijders, T. A. B., & Bosker, R. J., (Eds.). (2016). *Multilevel network analysis for the social sciences: Theory, methods and applications*. SAGE.
- 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>
- Soral, W., Malinowska, K., & Bilewicz, M. (2022). The role of empathy in reducing hate speech proliferation. Two contact-based interventions in online and off-line settings. *Peace and Conflict: Journal of Peace Psychology*, 28, 361–371. <https://doi.org/10.1037/pac0000602>
- Teerenstra, S., Lu, B., Preisser, J. S., van Achterberg, T., & Borm, G. F. (2010). Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics*, 66(4), 1230–1237. <https://doi.org/10.1111/j.1541-0420.2009.01374.x>
- Volk, A. A., Provenzano, D. A., Farrell, A. H., Dane, A. V., & Shulman, E. P. (2021). Personality and bullying: Pathways to adolescent social dominance. *Current Psychology*, 40(5), 2415–2426. <https://doi.org/10.1007/s12144-019-00182-4>
- Wachs, S., Bilz, L., Wettstein, A., Wright, M. F., Kansok-Dusche, J., Krause, N., & Ballaschk, C. (2022). Associations between witnessing and perpetrating online hate speech among adolescents: Testing moderation effects of moral disengagement and empathy. *Psychology of Violence*, 12(6), 371–381. <https://doi.org/10.1037/vio0000422>
- Wachs, S., Bilz, L., Wettstein, A., Wright, M. F., Krause, N., Ballaschk, C., & Kansok-Dusche, J. (2022). The online hate speech cycle of violence: Moderating effects of moral disengagement and empathy in the victim-to-perpetrator relationship. *Cyberpsychology, Behavior and Social Networking*, 25, 223–229. <https://doi.org/10.1089/cyber.2021.0159>
- Wachs, S., Gámez-Guadix, M., & Wright, M. F. (2022). Online hate speech victimization and depressive symptoms among adolescents: The protective role of resilience. *Cyberpsychology, Behavior and Social Networking*, 25(7), 416–423. <https://doi.org/10.1089/cyber.2022.0009>
- Wachs, S., Krause, N., Wright, M. F., & Gámez-Guadix, M. (2023). Effects of the prevention program “HateLess. Together against hatred” on adolescents' empathy, self-efficacy, and countering hate speech. *Journal of Youth and Adolescence*, 52, 1115–1128. <https://doi.org/10.1007/s10964-023-01753-2>
- Wachs, S., Mazzone, A., Milosevic, T., Wright, M. F., Blaya, C., Gámez-Guadix, M., & O'Higgins Norman, J. (2021). Online correlates of cyberhate involvement among young people from ten European countries: An application of the routine activity and problem behaviour theory. *Computers in Human Behavior*, 123, 106872. <https://doi.org/10.1016/j.chb.2021.106872>
- Wettstein, A. (2021). Hate speech. Aggressionstheoretische und sozialpsychologische Erklärungsansätze. In S. Wachs, B. Koch-Priewe, & A. Zick (Eds.), *Hate speech—Multidisziplinäre Analysen und Handlungsoptionen* (pp. 227–251). Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-31793-5\\_10](https://doi.org/10.1007/978-3-658-31793-5_10)
- Wilson, M. S., & Liu, J. H. (2003). Social dominance orientation and gender: The moderating role of gender identity. *British Journal of Social Psychology*, 42, 187–198. <https://doi.org/10.1348/014466603322127175>
- Yates, F., & Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society: Series B (Methodological)*, 15(2), 253–261. <https://doi.org/10.1111/j.2517-6161.1953.tb00140.x>
- Zubielewicz, E., Osborne, D., Milojev, P., & Sibley, C. G. (2023). Social dominance orientation and right-wing authoritarianism across the adult lifespan: An examination of aging and cohort effects. *Journal of Personality and Social Psychology*, 124(3), 544–566. <https://doi.org/10.1037/pspi0000400>
- Zych, I., Ttofi, M. M., & Farrington, D. P. (2019). Empathy and callous-Unemotional traits in different bullying roles: A systematic review and meta-analysis. *Trauma, Violence & Abuse*, 20(1), 3–21. <https://doi.org/10.1177/1524838016683456>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Castellanos, M., Wettstein, A., Wachs, S., & Bilz, L. (2023). Direct and indirect effects of social dominance orientation on hate speech perpetration via empathy and moral disengagement among adolescents: A multilevel mediation model. *Aggressive Behavior*, 1–11. <https://doi.org/10.1002/ab.22100>