MUNI FSS

JOURNAL OF PSYCHOSOCIAL RESEARCH ON CYBERSPACE
CYBERPSYCHOLOGY

Ahrens, E., Bojarskich, V., & Rothmund, T. (2026). Towards a typology of online users in response to incivility: A latent profile analysis across six countries. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 20*(3), Article 1. https://doi.org/10.5817/CP2026-3-1

Towards a Typology of Online Users in Response to Incivility: A Latent Profile Analysis Across Six Countries

Emily Ahrens, Vladimir Bojarskich, & Tobias Rothmund

Institute of Communication Science, Friedrich Schiller University Jena, Germany

Abstract

Incivility is pervasive in the online communication landscape and has been linked to harmful consequences for individuals and the quality of public discourse. Understanding how people engage with online incivility is crucial for identifying mechanisms that may reinforce or mitigate such interactions. Although previous research investigated how incivility is experienced in different social roles such as victims, bystanders and perpetrators, most studies have examined these roles separately rather than analyzing how experiences, behaviors and attitudes toward incivility co-occur within individuals. To systematically identify and distinguish patterns of behaviors, experiences and attitudes associated with incivility, a Latent Profile Analysis was carried out, using a large-scale data set (N = 4,555) surveying young adults aged 18 to 26 (M = 21.69, SD = 2.24) from six countries (US, UK, Spain, Poland, France, Finland). Our results indicate that five profiles or types of users can be distinguished. Individuals are categorized as: perpetrator-victims; uninvolved; observers; defenders; tolerators. Additional analyses suggest that moral emotions vary systematically between these profiles. While perpetrator-victims report more pride, hate and guilt emotions when encountering incivility than any other profile, defenders indicate higher degrees of shame. By identifying distinct patterns of involvement in online incivility and their associated emotional responses, the findings provide insights into the psychological dynamics underlying uncivil online interactions and highlight potential targets for preventive interventions and future research.

Keywords: incivility; hate speech; flaming; bystander intervention; moral emotions; latent profile analysis

Editorial Record

First submission received:
October 18, 2024

Revisions received:
August 21, 2025
March 13, 2026

Accepted for publication:
May 8, 2026

Editor in charge:
Lenka Dedkova

Introduction

Ever since digital media facilitated new routes of communication, there has been uncivil discourse online (Papacharissi, 2004). Yet, especially adolescents and young adults appear likely to encounter and experience online hate, both as witnesses and as targets (Kumar, 2025; Wang et al., 2019). Research suggests that exposure to online incivility triggers negative emotions and thereby reduces satisfaction and willingness to participate in online discussions (Gervais, 2015; Lück & Nardi, 2019). Furthermore, incivility may also lead to hostile cognitions (Rösner et al., 2016) or negative emotions that facilitate retaliation (Chen & Lu, 2017). Additionally, incivility is suspected to increase polarized judgements in discussions (Anderson et al., 2013) and generally disrupt

deliberative online discourse (Lück & Nardi, 2019). In its most extreme form, online incivility has been associated with real-life violent hate crimes (Müller & Schwarz, 2020).

These negative effects of incivility online, combined with its continuously prominent role in public online discourse (Salgado et al., 2023; Sun et al. 2021) point to the necessity of a better understanding of intra- and interpersonal psychological mechanisms that are involved when people are exposed to online incivility. Prior research has already identified important predictors of individual experiences and behaviors in different roles, such as perpetrators, victims, or bystanders—reflecting the sender, receiver, and observer roles in communication theory (see for example Shannon, 1948). For example, findings suggest that being more frequently exposed to uncivil content might increase the risk of becoming victim and perpetrator of incivility (Frischlich et al., 2021; Wachs et al., 2019). However, much of this work relies on variable-centered approaches, which focus on relationships between isolated variables across an entire sample. While useful, this approach treats individuals as homogenous units, potentially obscuring meaningful patterns within individuals. For instance, it cannot identify whether the same individuals who frequently perpetrate online incivility also frequently experience victimization - since it analyzes these variables independently rather than jointly within individuals. To address this limitation, the current study adopts a person-centered approach using Latent Profile Analysis (LPA), which allows us to classify individuals based on shared patterns across indicators. By incorporating measures of exposure to incivility, personal victimization, perpetration, support for retaliation, and defending behavior, we aim to take a broader look at individual involvement in uncivil online discourse. We then compare these profiles across six countries, alongside demographic factors and moral-emotional responses. Summarizing our research goal, we aim to investigate how experiences of and engagement with online incivility are systematically linked within individuals, analyze the intraindividual dynamics involving emotional reactions and potentially derive insight into pathways for tailored interventions and prevention strategies.

Online Incivility

Incivility reflects violations of general communicative norms, such as insults that can be directed against individuals or groups (Coe et al. 2014). This conceptualization of incivility can be used as an umbrella term for a broader range of negative or insulting speech, such as hate speech and flaming (e.g., Aiken & Waller, 2000; Izquierdo-Montero et al., 2022; Thompsen & Foulger, 1996; Warner & Hirschberg, 2012). Accordingly, when referring to prior research or specific survey measures, we use the terminology employed in those studies (e.g., hate speech or uncivil discourse), while treating these related concepts as manifestations of online incivility.

The emergence of social media platforms and their growing role in election campaigns, protests, and political discourse—particularly since the late 2000s—raised public and scholarly concerns about increasing levels of uncivil and hateful speech. Empirical research, however, suggests that overall levels of incivility on major platforms such as Reddit and Twitter (now X) have remained relatively stable over time (Sun et al., 2021; Theocharis et al., 2020). At the same time, research shows that levels of incivility can fluctuate in response to major political events and platform governance, with spikes often occurring around salient moments such as elections or crises and declines following content moderation practices such as deplatforming (Jhaver et al., 2021; Rasmussen & Petersen, 2023; Sun et al., 2021 Theocharis et al., 2020; Saleem & Ruths, 2018).

While current evidence provides little support for a sustained overall increase in online incivility, emerging research suggests that incivility may be concentrated among particular users. Studies using user- and person-centered approaches show that some individuals repeatedly engage in uncivil behavior across online communities, including and beyond political contexts, and that higher levels of online incivility are associated with adverse life experiences such as criminal justice involvement or time spent in foster care (Mamakos & Finkel, 2023; Rasmussen et al., 2024). These findings highlight the importance of person-centered approaches that identify how different forms of engagement with incivility—such as perpetration, victimization, and bystander behavior—co-occur within individuals. Yet research examining how different forms of engagement with incivility—beyond the common separate focus on perpetrators or victims—cluster within individuals remains limited.

Online Encounters With Incivility

People engage with online incivility in different psychological modalities (e.g., cognitive, affective, behavioral) and in different social roles (e.g., perpetrator, victim, observer). In regard to experiential involvement, we distinguish between (a) *exposure* to hateful content in terms of the general quantity of encounters with incivility from any

possible social perspective, and (b) *victimization*, which refers to instances in which a person experienced themselves as the target of incivility. In addition, we consider (c) *retaliation acceptance,* indicating the normative legitimization of reactive forms of aggression against a perpetrator, and thus captures individuals' cognitive-evaluative responses to incivility. When it comes to behavioral engagement with incivility, distinctions are made between (d) *perpetration*, whether people use uncivil speech in an online context themselves and (e) *bystander behavior* in the form of intervening to prevent or counteract incivility.

There is generally much research on predictors and consequences of each of these different forms of experiences and behavioral engagements with online incivility. Madriaza et al. (2025) provided a meta-analytic review of studies on exposure to hate speech and found that exposure negatively affects individuals' psychological well-being. Ng et al. (2020) systemized dispositional (e.g., attentional impulsivity), developmental (e.g., educational level) and social (e.g., time spent online) predictors of perpetration behavior (see also Frischlich et al., 2021). There is consistent evidence that victimization experiences are likely to elicit depressive symptoms (e.g., Wachs et al., 2022) and there is also research on boundary conditions that make bystander intervention more or less likely (e.g., Obermaier et al. 2023).

In this paper, we focus on the co-occurrence of different forms of experiences and behavioral engagements in different social roles with online incivility. We believe that this perspective complements existing research by allowing a better understanding of the intra- and interpersonal dynamics associated with incidents of online incivility. There is already some research on these relations. For example, *perpetration* of online incivility seems to be systematically linked with *exposure* to incivility in general (Gervais, 2015; Hsueh et al., 2015; Wachs et al., 2019) and *victimization* experiences in specific (Festl, 2014; Frischlich et al., 2021; Wachs et al., 2021). These findings suggest a cycle of violence (Wachs et al., 2022) that reflects a psychological or social link between experiencing online incivility from a victim or observer perspective and the engagement in this behavior as a perpetrator. There are different theoretical explanations for this link. Bilewicz and Soral (2020) argue that exposure to online incivility might desensitize individuals, reduce empathic reactions and erode social norms that prevent people from engaging in incivility as a perpetrator. The same authors suggest that emotional reactions to incivility, such as anger and contempt, might facilitate active engagement in online incivility (Bilewicz et al., 2025).

Within this cycle of violence, it is also reasonable to assume that perpetrators are more likely to be victimized due to retaliation from victims or bystanders. A key mechanism that may shape these dynamics are legitimizing evaluations of harmful behavior, that is, cognitive processes through which individuals justify or accept aggressive responses in situations of incivility. One such evaluation is retaliation acceptance, which reflects the normative legitimization of responding to incivility with reactive aggression. For instance, Festl et al. (2014) report a positive association between retaliation behavior and victimization. Related research also suggests that bystander intervention may depend on such legitimizing evaluations. For instance, *bystander intervention* is, to some extent more likely the more people report exposure to online hate (Hansen et al., 2024; Obermaier, 2022) or victimization experiences (Hansen et al., 2024; Henson et al., 2019). Moxey and Bussey (2019) found that higher levels of moral disengagement, another cognitive mechanism that facilitates the justification of harmful behavior were associated with aggressive bystander intervention, while constructive responses such as comforting the victim were predicted by lower moral disengagement. Taken together, these findings suggest that perpetrators may themselves have prior victimization experiences, while victims and bystanders may also endorse retaliatory justifications. Yet research rarely examines how these dynamics co-occur within the same individuals.

While there are person-centered analyses on incivility related concepts like predictors of victimization (Obermaier & Schmuck, 2022) and different offline bystander behaviors (Wachs et al., 2024), the current analysis takes a broader look at multiple behaviors and experiences related to online incivility. The goal of person-centered approaches is to identify distinct subgroups which report similar patterns of experiences and behaviors (Bauer, 2022). In line with previous works, this study will make use of a Latent Profile Analysis and analyze additional descriptive variables for each resulting profile. By doing so, we are able to identify distinct response patterns and estimate the relative frequency of these patterns within the population.

Multigroup Latent Profile Analysis

A Latent Profile Analysis allows for the investigation of the interrelations of multiple variables across individuals. Research shows that people can be categorized into different profiles of individuals using latent class or profile analyses, for example, based on their attitudes towards climate change (Rhead et al., 2018) or the COVID-19 pandemic (Rothmund et al., 2022). In the context of online incivility and related behaviors, person-centered

approaches have been applied to examine patterns of victimization and its predictors (Obermaier & Schmuck, 2022) as well as different forms of bystander behavior in peer contexts (Wachs et al., 2024). Building on this line of research, the present study extends the person-centered perspective to multiple behaviors and experiences related to online incivility. Identifying groups of people with similar patterns of experiences or behaviors related to an issue can be useful for prevention research (Nylund-Gibson & Hart, 2014), as previously demonstrated for instance in relation to students at risk for depression and anxiety (Marcotte & Starrs, 2023) or predictors of substance use (Cleveland et al., 2010). The general benefit here is that these groups can be targeted with specific interventions that are tailored towards the specific intraindividual dynamics that are relevant for these individuals.

In line with previous applications of latent class or profile analysis (Kácha et al. 2022), the current study will use data from multiple countries to investigate the comparability and generalizability of latent groups across samples. The dataset that is used for this study includes data from the US, UK, Spain, Poland, France and Finland. By comparing samples from countries that are relatively similar in internet use (World Bank Group, 2024) and political regulation of online content (Simmons, 2022), this approach allows us to evaluate the generalizability of results across comparable samples. Despite similarities, there is reason to expect variations in the online environment and associated experiences between countries. Previous research on problematic internet use has shown that for instance, the degree to which internet use is associated with problematic behaviors can vary between otherwise similar countries (Laconi et al., 2018). A report on children's experiences with cyberhate across several EU countries reveals, that there are country specific differences regarding cyberhate exposure, victimization and aggression. For instance, exposure as well as victimization and aggression in relation to cyberhate are higher in Poland than France (Machackova et al., 2020). From a linguistic perspective, a comparison of uncivil content in Polish and English shows that the perception and expression of incivility may differ according to cultural differences (Lewandowska-Tomaszczyk, 2017). These findings suggest that there are country specific differences in levels of online incivility. Additionally, cultural norms, language practices, and media environments can shape how incivility is perceived, tolerated, or responded to, and thus provide the theoretical basis for expecting some degree of cross-national variation.

Two research questions are posed, to investigate latent profiles related to incivility and their consistency across several countries.

RQ1: Which latent groups of people can be identified based on their individual experiences and behavioral engagement with online incivility?

RQ2: How are latent groups of people similar or different across countries?

Demographics and Emotions

To provide more detailed analyses of profiles, additional variables can be used to describe characteristics of individuals that are grouped together based on the profile criteria. For this purpose, demographic variables and emotional reactions, more specifically emotions in reaction to remembering uncivil content, will be examined for the respective groups in this study.

Considering demographic data, men tend to participate more in incivility (Costello & Hawdon, 2018; Moxey & Bussey, 2019; Šincek, 2021). However, evidence on interventive behaviors is mixed. Depending on whether intervention is expressed by reporting uncivil content or related to aggressive behaviors, it is associated more with female or male gender respectively (DiFranzo et al., 2018; Moxey & Bussey, 2019). As the aim of the current study is to map out patterns of behavior, attitudes and experiences across all individuals, gender will be considered across all emerging groups.

Understanding emotional reactions to incivility can provide further insight into the intraindividual dynamics that link different experiences and behaviors towards online incivility. As outlined, there is consistent evidence that exposure to incivility is associated with anger and contempt, which then can motivate bystander intervention or even active engagement in online incivility (e.g., Bilewicz et al., 2025; Gervais, 2019; Hansen et al., 2024; Lück & Nardi, 2019; Masullo et al., 2020). However, other moral emotions are generally rather overlooked. For example, self-conscious moral emotions such as pride, shame and guilt are important indicators of moral evaluations and are likely to impact whether and how people legitimize online incivility or behavioral reactions to it (Tangney & Fischer, 1995). Although these self-conscious emotions are most relevant for evaluating one's own behavior (Tracy & Robins, 2004), they can also be felt when appraising the actions of others, especially members of an ingroup (Lickel et al., 2011). Miranda et al. (2020) for instance find that observing workplace incivility induces

shame and guilt. Whereas shame is generally associated with avoidance behaviors, guilt should be more likely to motivate moral compensation, such as retaliation against the perpetrator or support of victims (e.g., Rothmund & Baumert, 2013). While pride is often involved in responses to positive events and generally facilitates prosocial behavior, group-based pride can be associated with antisocial behavior and motivate aggressive acts that are motivated by a desire to affirm or protect the group's identity and status (Sullivan & Day, 2019). Thus, pride may not always function as a socially constructive emotion; its behavioral implications depend on the social context and the object of pride. However, as our measure of pride is not only referring to instances of perpetration, we cannot determine which specific form of pride may be activated in this context. Given that participants reported their emotions in response to witnessing uncivil content, pride may also arise in the context of observing or being affected by such interactions.

We aim to analyze different emotional reactions to incivility as indicators of intraindividual dynamics that are potentially involved in specific profiles. In order to do that, we compare the frequency of experiencing and expressing specific moral emotions between latent profiles. Understanding which groups are more likely to respond to incivility with different emotions could not only inform us about the underlying intraindividual processes, but also provide a theoretical foundation for tailored interventions, enabling platforms and educators to use communicative messages in order to stimulate processes of emotion regulation (for an overview, see Bettis et al., 2022). Therefore, we pose a third research question.

RQ3: How do people from the latent groups differ regarding demographics and emotional reactions toward incivility?

Present Research

The general goal of this study is to identify and to analyze distinct patterns of responses towards online incivility within individuals, and to better understand intra- and interindividual dynamics that are involved in for different groups of people. To answer our research questions, we used survey data from the US, UK, Spain, Poland, France and Finland of young adults between the ages 18-26, to conduct a Latent Profile Analysis. Young adults are a particularly relevant age group, as they are frequently exposed to online incivility and appear to be especially vulnerable to the effects of negative online interactions (Kumar, 2025; Politte-Corn et al., 2023; Wang et al., 2019). The input variables for the LPA (perpetration, victimization, exposure to incivility, bystander behavior and retaliation acceptance) as well as the additional descriptive variables (sex, anger, hate, pride, shame and guilt) are calculated from items of a self-report online survey. Based on the input variables, we identify latent groups of people regarding their incivility related experiences and behaviors and compare the LPA generated groups across national samples. After identifying the optimal latent profile solution, individuals were assigned to profiles based on their most likely class membership. These assignments were then used for descriptive and comparative analyses. Finally, the descriptive variables were used to compare the latent groups based on their expression of different emotions in response to incivility and sex differences.

Methods

Sample

The data used for this study was collected by Hawdon et al. (2018). The group of researchers made use of a Survey Sampling International (SSI) panel to collect data on the topic of incivility online, including items on experiences of victimization, perpetration, attitudes toward hateful content, bystander behavior as well as political attitudes and demographic information. SSI recruits panel members through multiple channels, such as telephone-based recruitment, online advertisements, and other opt-in methods. Individuals who join the panel typically complete an initial screening process, and panel participation is supported through incentive systems designed to encourage attentive and consistent participation. Respondents for the present survey were randomly selected from the panel members in the six countries included in the study. After excluding 508 missing cases in preparation for the LPA, the sample for the current study consisted of N = 4,555 young adults. Comparisons between included and excluded participants revealed statistically significant, small to moderate differences for some variables (see Supplementary Material S5 for detailed results). Participants were aged 18 to 26 (M = 21.69, SD = 2.24) from the US, UK, Spain, Poland, France and Finland. Two thirds (67.72%) of respondents were women and 52,53% of participants reported to spend six or more hours online daily.

Data Collection

The survey was conducted online in May 2018 across the six countries. The content of all surveys was identical and translated from English to each country's majority language if necessary. The translation was carried out and additionally reviewed by native speakers of the respective languages. Participants received information about the purpose of the study and anonymization of their information as well as the option to withdraw.

Measures

Item formulations for each of the following measures can be found in Supplementary Material S4.

Perpetration

Perpetration was measured using two self-report items that captured different aspects of participants' engagement in harmful online behavior. The first item asked: *Have you ever produced online material that other people would likely interpret as hateful or degrading?* with a binary response format (yes, no). The second item assessed the frequency of joining in harmful online interactions: *When people on social networking sites are being mean or offensive, how often does each of the following happen? - You join the mean behavior*, answered on a 4-point scale (never to frequently). The two items were moderately associated (Spearman's $\rho = .37$). To derive a perpetration indicator for each participant, responses were recoded to reflect presence (1) or absence (0) of perpetration behavior on each item and then summed into a composite score ranging from 0 (*no indication of perpetration on either item*) to 2 (*indications on both*).

Victimization

Victimization was measured by asking participants how frequently they were attacked in the last three months because of ethnicity or race, nationality, sexual orientation, religious conviction/belief, disability, the person's sex or gender, or sexual orientation (Cronbach's $\alpha = .90$). All seven items were measured on a scale from 1 (*never*), 2 (*once or twice*), 3 (*3 to 5 times*), 4 (*6 to 10 times*), to 5 (*11 or more times*). Scores were counted as missing values when less than half of the items were answered, otherwise, an average score was calculated from the remaining responses.

Bystander Intervention

Participants were also asked to answer on a scale from 1 (*never*) to 4 (*frequently*), how frequently they reacted to mean or offensive content by defending *the person or group that is being attacked*. Bystander intervention was measured with a single item.

Retaliation Acceptance

Three items from a series of questions about hate speech acceptance were combined, all relating to the acceptability of producing hateful content in response to an attack on oneself or others (*It is OK to send hateful or degrading messages against someone online if they start to attack you first*, *It is OK to send hateful or degrading messages against someone online if they insult your friends or your family*, *It is OK to send hateful or degrading messages against someone online if these people make fun of you or insult you because of your religion, your origin or the color of your skin*). Cronbach's alpha was $\alpha = .88$. Participants' responses were recorded on a scale from 1 (strongly agree) to 5 (strongly disagree).

Hate Exposure

To measure hate exposure, one item was used where participants were asked to indicate how frequently they were exposed to uncivil content online (*How frequently do you see material online that expresses negative views toward some group?*) on a scale from 1 (*never*) to 4 (*frequently*).

Emotional Reactions

In addition to the five input variables for the latent profile analysis, perpetration, victimization, retaliations acceptance, exposure and defending victims, emotional reactions to incivility were measured to further describe the emerging profiles. The emotions in reaction to incivility were recorded by asking participants how seeing uncivil content in the past made them feel. Each of the five emotional reactions (angry, hateful, guilty, proud and ashamed) could be rated on a scale from 1 to 100.

Results

Analysis Strategy

In the first part of the analysis, we identified a latent profile solution across countries and for each country individually, grouping participants according to their similarities across input variables using the R package *mclust* (Scrucca et al., 2023). The models were estimated as Gaussian mixture models assuming normally distributed continuous indicators within profiles and were fitted using maximum likelihood estimation via the Expectation–Maximization algorithm implemented in the *mclust* package. Prior to profile estimation, all input variables were z-standardized ($M = 0$, $SD = 1$) to ensure comparability across differently scaled measures.

Identification of Profiles

To identify the optimal number of profiles, the Bayesian information criterion (BIC) and Akaike information criterion (AIC) were calculated for the complete dataset and each national sample individually for 1-9 latent profiles. The upper limit of nine profiles was chosen because solutions beyond this point showed only marginal improvements in model fit while substantially reducing interpretability, and restricting the range ensured a parsimonious set of candidate models. In addition to statistical fit indices, the selection of the final solution also considered the theoretical interpretability of the resulting profiles. When using the data from all countries combined ($N = 4{,}555$), the entropy value for this solution was 0.85, indicating good classification accuracy and clear separation between profiles. The lowest BIC and AIC values indicated a five-profile solution as the best fit (see Table 1). The same pattern was observed for the national samples of the UK and France. Regarding the sample from the US, Spain and Finland, more than five profiles were associated with lower BIC and AIC values. However, in line with Nylund-Gibson and Choi (2018), an elbow point indicated that a five-profile solution sufficiently fits the data for each of these samples. Elbow points refer the points on a plot, where the improvement in model fit starts to level off, indicating an optimal balance between model complexity and explanatory power. Only the sample from Poland did not have a preferential fit of five profiles. In order to ensure comparability of results across countries, a five-profile solution was applied to this sample as well. The according fit indices per country, as well as detailed descriptions of the country-specific profiles can be found in the supplemental material.

Overall, the five identified latent profiles showed consistent patterns of behaviors, and experiences, which differed substantially between the groups (Figure 1). In the following sections, we describe the profiles using z-standardized means. These scores indicate each profile's deviation from the overall sample mean: a value of 0 corresponds to the sample average across all included cases, while a value of 1 represents a mean one standard deviation above that average. Importantly, these values describe relative differences within the sample rather than absolute levels of behavior or experience.

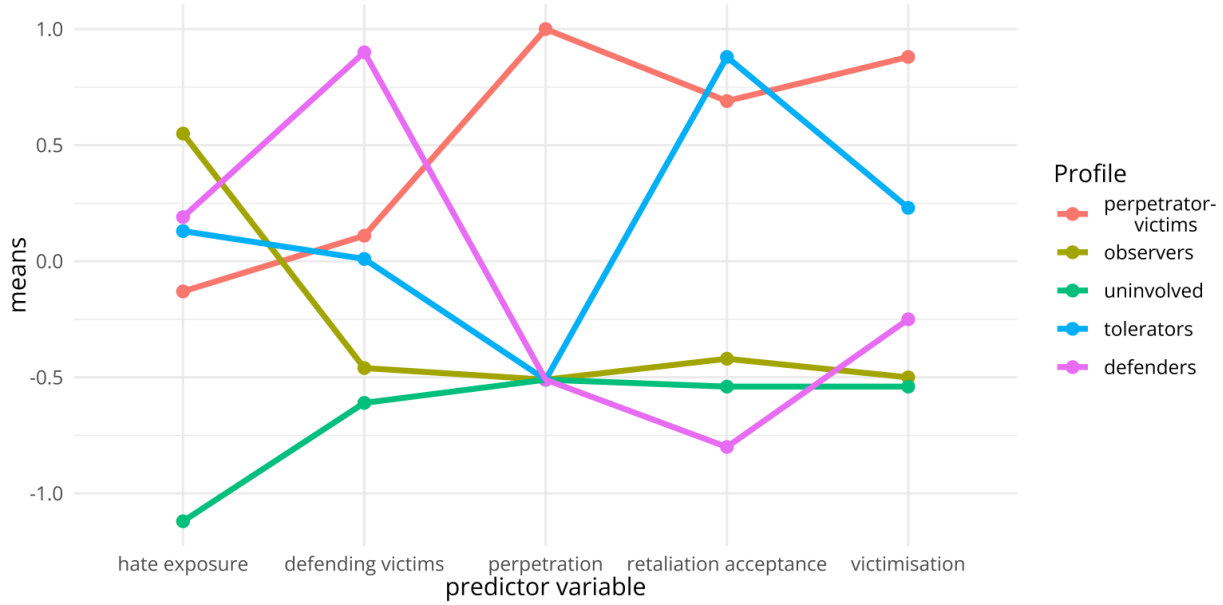
Across all six countries, similar patterns of profiles were observed. With one exception (i.e. no Spanish *observer* profile, see paragraphs *Observers* and *Defenders*), the same profiles were found for all six countries. The proportions of profile sizes, as detailed in Table 2, are often comparable but vary across samples (see supplemental materials for country-specific results).

Table 1. *LPA Fit Indices AIC and BIC Across Numbers of Profiles Based on the Full Dataset.*

N	AIC	BIC
1	61911.38	62039.86
2	51771.88	52009.57
3	47007.46	47354.35
4	44274.93	44731.03
5	**41321.64**	**41886.95**
6	47747.67	48390.07
7	48283.69	49028.87
8	57102.53	57500.82
9	57113.72	57550.55

Note. Lowest scores are marked in bold.

Figure 1. *Five Latent Profiles Based on the Full Dataset.*



Note. Figure displays z-standardized means.

Table 2. *Sample Sizes and Per-Country Distribution.*

	Overall (*n* = 4,555)	US (*n* = 1,000)	UK (*n* = 879)	Spain (*n* = 680)	Poland (*n* = 691)	France (*n* = 844)	Finland (*n* = 461)
Perpetrator-victims	1,046 (23.0%)	305 (30.5%)	215 (24.5%)	144 (21.2%)	50 (7.2%)	198 (23.5%)	144 (31.2%)
Observers	1,074 (23.6%)	148 (14.8%)	139 (15.8%)	—	112 (16.2%)	163 (19.3%)	93 (20.2%)
Uninvolved	653 (14.3%)	93 (9.3%)	84 (9,6%)	147 (21.6%)	59 (8.5%)	117 (13.9%)	67 (14.5%)
Tolerators	878 (19.3%)	181 (18.1%)	300 (34.1%)	91 (13.4%)	366 (53.0%)	149 (17.7%)	127 (27.5%)
Defenders	904 (19.8%)	273 (27.3%)	141 (16.0%)	205 (30.1%)	104 (15.1%)	217 (25.7%)	30 (6.5%)
Defender-victims	—	—	—	93 (13.7%)	—	—	—

Note. Percentages indicate the relative size of profiles within the (country-specific) sample. The *Overall* column is based on the latent profile analysis of the pooled sample and is not calculated by summing the country-specific counts.

Perpetrator-Victims

The group of *perpetrator-victims* stood out based on their uniquely high relative scores on both perpetration and victimization, as well as high retaliation acceptance. Scores on defending victims and hate exposure were close to the overall sample mean. Notably, while all other profiles showed below-average perpetration probabilities,

perpetrator-victims indicated prior perpetration at approximately +1 *SD* above the mean across all countries. This profile was relatively large in most national samples (21–31% of respondents), but much smaller in Poland (7.2%), where the *tolerator* profile also had perpetration scores near the sample mean. Victimization was highest for perpetrator-victims in all countries except Spain, where the *defender-victim* profile had a slightly higher score. Retaliation acceptance was consistently elevated – ranking highest among profiles in the US, UK, Poland, and Finland, and second-highest in Spain and France.

Uninvolved

The *uninvolved* consisted of people who, compared to other profiles, reported hate exposure well below the overall sample mean and scored below average on all other variables (defending victims, perpetration, acceptance of retaliation and victimization). In the Finnish sample, individuals with an *uninvolved* profile showed an average score for defending victims. Similarly, in the Spanish sample, hate exposure was slightly higher than in the other national *uninvolved* profiles. Overall, this profile showed lowest relative involvement of any kind in uncivil discourse compared to other profiles.

Observers

This profile scored relatively low on perpetration, victimization, acceptance of incivility, and defending victims. What distinguished *observers* from the *uninvolved* profile was their relatively high exposure to online incivility. This profile did not emerge in the Spanish sample; instead, Spain appeared to have two variations of the *defender* profile with average and above-average scores on exposure.

Defenders

The *defenders* defining feature was the relatively high score on defending victims, while scores on perpetration and acceptance of retaliation were below-average. Hate exposure varied between countries from average to above-average scores. In the Spanish sample, a second variation of this profile emerged: the *defender-victim* profile. Like the standard *defender* profile, it showed a high score on defending victims, but it also exhibited relatively high victimization and hate exposure. This pattern was similar to the *defender* profiles in the UK and France, which also showed elevated victimization relative to other national profiles. Overall, *defenders* consistently scored high on defending victims, while their victimization experiences varied by country, ranging from below to above the overall sample mean.

Tolerators

In all national samples, a *tolerator* profile emerged, characterized by relatively high acceptance of retaliation and average scores on exposure to hate, defending victims and victimization experiences. Compared to other profiles, this group showed most volatility—in size (13,4% in Spain, and 53% in Poland) as well as in country-specific patterns. Notably, individuals from this group scored second highest on victimization compared to other profiles. Overall, *tolerators* showed moderate and volatile involvement in uncivil discourse with an above-average acceptance to retaliate.

Description of Profiles: Demographics and Emotions

In addition to the predictor variables of the LPA, we investigated the association of the resulting profiles and emotions in relation to online incivility. Demographic data was also considered to describe profiles and possible differences between them. Differences between profiles in continuous variables, such as emotional responses, were examined using one-way analyses of variance (ANOVA) with Tukey's Honestly Significant Difference (HSD) tests used for post hoc pairwise comparisons. Effect sizes for pairwise differences are reported as Cohen's *d*. The subsequent section will provide insights into the variations among profiles regarding demographic data and expressed emotions in relation to incivility (see Table 3). More detailed data on the country-specific differences can be accessed in the supplemental material.

Demographics

The distribution of sex across the five groups was examined using Pearson's Chi-squared test. As mentioned above, the gender distribution in the overall sample was not even, women made up two thirds of the sample. The analysis revealed a significant association between sex and group membership, $\chi^2(4,4542) = 204.32$, $p < .001$, Cramér's $V = .21$. A post hoc pairwise proportion tests with Bonferroni correction revealed, that *perpetrator-victims* are significantly more likely to be male compared to the other profiles (47.99% male), while *observers* (24.28% male) and *defenders* (21.06% male) were more likely to be female compared to all other groups.

Table 3. *Mean Values, Standard Deviations and Significant Differences Across Profiles in the Overall Sample.*

	Overall	Perpetrator-victims (22.96%)	Observers (23.58%)	Uninvolved (14.34%)	Tolerators (19.28%)	Defenders (19.85%)	Test statistic (*df*)
Sex	32.28% male	47.99% male[2,3,4,5]	24.28% male[1,3,4]	32.00% male [1,2,5]	35.09% male[1,2,5]	21.06% male[1,3,4]	204.32 (4,4542)
Age	21.7 (2.24)	21.8 (2.28)[4]	21.7 (2.15)[4]	21.9 (2.29)[4]	21.3 (2.23)[1,2,3,5]	21.8 (2.23)[4]	8.64 (4,4540)
Education	3.61 (1.91)	3.47 (1.83)[2,5]	3.73 (2.00)[14]	3.67 (1.97)	3.48 (1.87)[2]	3.72 (1.85)[1]	4.36 (4,453)
Religiosity	2.24 (1.33)	2.49 (1.31)[2,3,4,5]	2.10 (1.27)[14]	2.07 (1.31)[1,4]	2.27 (1.34)[1,2,3]	2.23 (1.37)[1]	15.71 (4,4540)
Internet use	3.70 (1.29)	3.71 (1.34)	3.68 (1.22)	3.59 (1.32)[4]	3.82 (1.28)[3]	3.70 (1.28)	3.11 (4,4493)
Perpetration	0.30 (0.59)	1.29 (0.46)[2,3,4,5]	0 (0)[1]	0 (0)[1]	0 (0)[1]	0 (0)[1]	7077 (4,4550)
Exposure	2.72 (0.86)	2.60 (0.94)[2,3,4,5]	3.19 (0.40)[1,3,4,5]	1.76 (0.43)[1,2,4,5]	2.83 (0.87)[1,2,3]	2.86 (0.80)[1,2,3]	411.1 (4,4550)
Victimization	1.39 (0.66)	1.97 (0.94)[2,3,4,5]	1.06 (0.11)[1,4,5]	1.04 (0.09)[1,4,5]	1.54 (0.68)[1,2,3,5]	1.22 (0.26)[1,2,3,4]	485.4 (4.4550)
Retaliation acceptance	2.12 (1.27)	3.00 (1.39)[2,3,4,5]	1.60 (0.58)[1,3,4,5]	1.44 (0.52)[1,2,4,5]	3.26 (1.21)[1,2,3,5]	1.09 (0.19)[1,2,3,4]	1029 (4,4550)
Defending victims	2.42 (1.07)	2.54 (0.96)[2,3,5]	1.91 (0.93)[1,3,4,5]	1.73 (0.91)[1,2,4,5]	2.43 (1.11) [2,3,5]	3.37 (0.57)[1,2,3,4]	427.8 (4,4550)
Anger	67.17 (28.74)	59.14 (28.89)[2,4,5]	68.06 (27.97)[1,5]	63.89 (28.43)[5]	65.94 (30.55)[1,5]	76.94 (24.78)[1,2,3,4]	35.46 (4,2906)
Hate	41.09 (30.97)	48.18 (28.90)[2,3,4,5]	35.25 (29.70)[1,4,5]	36.05 (29.66)[1]	41.74 (30.96)[1,2]	40.47 (33.55)[1,2]	16.48 (4,2562)
Pride	13.99 (25.20)	31.8 (32.33)[2,3,4,5]	4.33 (11.93)[1,4]	7.46 (18.53)[1]	9.84 (21.07)[1,2]	4.61 (13.11)[1]	137 (4,1874)
Guilt	22.11 (27.19)	34.64 (30.78)[2,3,4,5]	14.08 (20.57)[1,4,5]	17.89 (22.45)[1]	19.17 (26.01)[1,2]	19.02 (25.98)[1,2]	51.82 (4,2146)
Shame	51.77 (34.31)	50.64 (31.40)[5]	49.23 (35.00)[5]	45.83 (33.63)[5]	51.33 (35.47)[5]	58.34 (35.06)[1,2,3,4]	7.92 (4,2610)

Note. Superscript numbers indicate $p < .05$ between according profiles: [1] = perpetrator victims, [2] = observers, [3] = uninvolved, [4] = defenders, [5] = tolerators.

Hate

Reported hatred differed significantly between profiles ($p < .001$) and was the highest in *perpetrator-victims*. Effects were of small to medium size compared to *observers* ($p < .001$, $d = 0.44$), *uninvolved* ($p < .001$, $d = 0.42$), *tolerators* ($p = .003$, $d = 0.22$) and *defenders* ($p < .001$, $d = 0.25$). Hate was lowest in the profile of *observers*, with significant small differences compared to *tolerators* ($p = .003$, $d = 0.21$) and *defenders* ($p = .026$, $d = 0.17$).

Pride

Profiles differed significantly in experienced and reported pride when encountering online incivility ($F(4,1874) = 137$, $p < .001$) and was highest in the groups of *perpetrator-victims*. Effects were of medium to large size compared to *defenders* ($p < .001$, $d = 1.17$), *observers* ($p < .001$, $d = 1.09$), *uninvolved* ($p < .001$, $d = 0.81$) and *tolerators* ($p < .001$, $d = 0.77$). In addition to these medium to large effect sizes, the difference between *observers* and *tolerators* was also significant with a small effect size ($p = .005$, $d = 0.34$).

Guilt

There were significant differences between the profiles with regard to their expression of guilt upon recalling incivility ($F(4,2146) = 51.82$, $p < .001$), with highest guilt in the group of *perpetrator-victims*. Effects were of medium to large size compared to *observers* ($p < .001$, $d = 0.78$), *uninvolved* ($p < .001$, $d = 0.58$), *tolerators* ($p < .001$, $d = 0.53$) and *defenders* ($p < .001$, $d = 0.54$). Guilt was lowest in the group of *observers* with additional significant small differences compared to *tolerators* ($p = .025$, $d = 0.22$) and *defenders* ($p = .026$ $d = 0.21$).

Shame

Responding to recalled incivility by indicating shame was also not equal across profiles ($F(4,2610) = 7.92$, $p < .001$). *Defenders*, with the highest mean score differed significantly from *uninvolved* ($p < .001$, $d = 0.36$), *observers* ($p < .001$, $d = 0.26$), *perpetrator-victims* ($p < .001$, $d = 0.23$), and *tolerators* ($p = .006$, $d = 0.20$), with generally small effect sizes.

Discussion

The general goal of this study was to identify and analyze distinct patterns of responses towards online incivility within individuals in order to better understand intra- and interindividual dynamics that are involved for different groups of people. Based on individuals' prior online experiences and behaviors - such as perpetration, victimization, bystander intervention, exposure to hateful content and retaliation acceptance - we identified five distinct latent profiles. First, the *perpetrator-victim* profile consists of people who reported the highest frequency of both, perpetrator behavior and victimization experience. This finding supports previous research indicating that perpetration of online incivility is systematically linked to victimization experiences within individuals (Festl, 2014; Frischlich et al., 2021; Wachs et al., 2021). Also consistent with previous findings, perpetrator-victims were more likely to be male than members of other profiles. However, our findings do not support the assumption that this group is also most likely to be exposed to online incivility in general (Gervais, 2015; Hsueh et al., 2015; Wachs et al., 2019). Second, the *defender* profile describes active bystanders who were the most likely to support victims but do not accept retaliation and are not being perpetrators themselves, while also reporting moderate experiences of online victimization. This group might be involved in non-violent forms of counter-speech or victim support in response to online incivility. Third, *tolerators* showed a relatively high acceptance for retaliation against perpetrators but below average participation in perpetration and tendency to defend victims. This group seems to be specifically characterized by their normative support for harsh reactions towards online perpetrators. Finally, the *uninvolved* profile and the *observer* profile describe individuals who were least likely to get actively involved in uncivil discourse compared to other profiles, although observers experience online incivility most frequently.

Congruence of Profiles Across Countries

The structures of the five profiles were largely consistent with exceptions across countries. In the Spanish sample, no *observer* group emerged, instead, there are two groups that can be classified as *defenders*. In the Polish sample, the *tolerator* group reports a higher frequency of perpetration compared to the other samples and seems to include more people that are categorized as perpetrator-victims in the other samples. These deviations highlight the importance of interpreting profile structures in light of national context. Besides these deviations, however, the similarities between the profiles for each national sample are striking (see Supplementary Material S2 for country specific visualizations of all profiles). Considering that cross-national differences have been found to play a role in the experience or expression of incivility (Laconi et al., 2018; Lewandowska-Tomaszczyk, 2017; Machackova et al., 2020), the similarity of profiles across countries indicates that the discovered profiles may be widely applicable to describe online environments.

The relative size of the profiles varies between national samples. Except for the Polish and Spanish sample that have specific profile characteristics (see above), two numbers stand out. First, it is striking that the group of uninvolved is very small in the US. This might reflect a higher general level of online incivility in the US compared to other countries that can be the result of a different legal situation (Rosenfeld, 2003). Second, the relative size of defenders is very low in Finland, which might reflect national differences in cultural norms but could also be a methodological artifact.

Emotional Reactions and Intraindividual Dynamics in the Perpetrator-Victim Profile

The *perpetrator-victim* profile is the group most actively involved in online incivility. While the co-occurrence of perpetration and victimization has been widely documented across contexts (Festl et al., 2014; Frischlich et al., 2021; Wachs et al., 2021; Wachs et al., 2022), examining how this co-occurrence relates to other incivility-related indicators at the individual level may help contextualize it. This profile also shows a high acceptance of retaliation, which supports previous assumptions about the dynamics of the perpetrator-victim relationship. Research on cyberbullying suggests that legitimizing evaluations of harmful behavior may play an important role in this

dynamic. For example, retaliation intentions and moral disengagement have been shown to moderate the victim–perpetrator association (Festl et al., 2014; Hood & Duffy, 2018). In this sense, victimization may be linked to aggressive retaliation if retaliation is morally accepted or if people legitimize retaliation and disengage from moral evaluations, for example by means of attributing blame to the victim—one of several strategies of moral disengagement. This aligns with the possibility of an escalation pathway that fuels a "cycle of violence" (Wachs et al., 2022), when victimhood combined with retaliation acceptance fuels aggressive responses. This interpretation is in line with the finding that perpetrator-victims reported the highest levels of hate and pride when exposed to online incivility.

At first, the fact that people belonging to the *perpetrator-victims* would report feelings of pride when remembering uncivil content seems puzzling. Pride is typically linked to morally virtuous acts rather than negative behavior (Haidt, 2003). However, morality is often a matter of perception. Perpetrators might still see themselves as morally virtuous because they do not view their actions as inherently wrong; instead, they may believe they engaged in incivility for a greater cause, such as protecting moral values (Van Zomeren et al., 2023). This could explain why retaliatory incivility, perceived as a justified response to prior provocation, might evoke feelings of pride (Osborne et al., 2025). Another explanation involves distinguishing between authentic and hubristic pride. Authentic pride is rooted in one's moral values and encourages prosocial behavior. In contrast, hubristic pride is based on external validation and can drive anti-social behavior (Krettenauer & Casey, 2015), often motivated by a desire to affirm or protect the group's identity and status (Sullivan & Day, 2019). Though the measurement of pride does not allow for a distinction of these forms, it is theoretically possible that when members of the *perpetrator-victim* group associate pride with uncivil behavior, they are experiencing hubristic pride, which enhances their group-based self-image. Thus, pride could be experienced as a group-focused emotion that serves the self in this context.

However, this group is also characterized by the highest levels of guilt as a specific moral reaction to offensive behavior. Guilt entails a negative evaluation of one's own behavior (Lickel et al., 2011) and is felt by individuals based on specific behaviors that they can control (Tracy & Robins, 2006). The self-conscious emotion is associated with a motive to repair (Lickel et al., 2005) and has been found in correlation to retaliation behavior and supporting victims (Miranda, 2020). Although the support of victims (measured as defending victims) is only moderate among *perpetrator-victims*, retaliation behavior and justification thereof are clear features of this profile. Therefore, guilt could appear as a self-focused emotion that serves to repair relationships with others. This finding indicates that perpetrator-victims might still be sensitive towards moral reflections about the cycle of violence.

Pride is associated with motives based on group identity while serving to enhance self-image, whereas guilt is self-focused and motivates behaviors that benefit others. Thus, these apparently contradictory emotions seem to serve different functions in processing one's own perpetration behavior and victimization experiences. Future research could further explore how pride and guilt interact in shaping individuals' responses to online incivility and how these emotions can be leveraged in the design of effective interventions. Importantly, the causal link between perpetrator behavior and experiences of victimization could also work in a reverse fashion, with victimization emerging as a consequence of prior perpetration. In this scenario, offensive behavior may provoke uncivil retaliatory behavior by others (see Bondü et al., 2016, for a discussion of such dynamics).

Emotional Reactions and Intraindividual Dynamics in the Defender Profile

Understanding the motives of *defenders* could give insight into the conditions under which people tend to engage in non-violent forms of bystander intervention when confronted with online incivility. Cross-national comparisons suggest that past victimization may motivate bystander intervention (Henson et al., 2019). However, victimization is not generally overrepresented in the *defender* profile. Additionally, low retaliation acceptance across all national samples indicates that defending victims of online incivility may not be related to moral disengagement. This aligns with the finding by Moxey and Bussey (2019) that moral disengagement is negatively correlated with non-violent bystander intervention in uncivil discourse. Thus, people who intervene in online discourse by defending victims do not seem to act based on aggression or a motive for revenge.

Interestingly, however, this group does seem to be motivated by anger, which is highest in this profile. This finding is in line with previous findings that anger motivates people to speak out against incivility (Masullo et al., 2020). However, shame, which has been linked to inaction in reaction to incivility (Miranda, 2020) is also highest among *defenders*. Even though the effect sizes for this association are rather small, understanding this seemingly paradoxical relationship may help to understand bystander motives. Literature on so-called vicarious shame, which describe experiencing emotions on behalf of someone else, might offer some insight into this finding (Lickel

et al., 2005). Vicarious shame can be experienced towards the behavior of a perpetrator, especially if the perpetrator is part of an ingroup with a shared identity. As opposed to a motive to repair as is the case for guilt, shame invokes a motive to distance. However, distancing, measured by Lickel et al. (2005) as not wanting to be associated with a person, does not have to imply inaction. Thus shame, though previously associated with inaction, may motivate intervention, when it is felt for another person, by creating a need to actively distance oneself from someone else's behavior.

Practical Implications and Future Research

Our research informs previous research on interventions about online incivility (for an overview see, Windisch et al., 2022). Based on the person-centered methodological approach, we identified specific individual profiles in the engagement with online incivility that reflect potential target groups for tailored interventions. Especially *perpetrator-victims* seem to be a well-suited group for specific communicative interventions. Furthermore, our research approach allows to inform the design of interventions based on the intra- and interpersonal dynamics that our analyses indicate. Our findings speak in favor of practical interventions that sensitize *perpetrator-victims* towards the pitfalls of moral disengagement as a driver of cycles of violence. Based on our results, such an approach is likely to include at least three elements: First, an educational approach that informs about the logic of reactive cycle of violence dynamics in interpersonal and intergroup communication settings (see Postmes et al., 2014). Second, a communicative format that sensitizes *perpetrator-victims* towards the pitfalls of moral disengagement as a driver of cycles of violence (see for example, Babaei, et al., 2020). Third, an emotion regulation approach that aims at developing skills in regulating hate and contempt when incivility is experienced online (see for example, Syrjämäki et al., 2022).

Additionally, the group of *defenders* might also be especially relevant for practical intervention research, given that their pattern seems to involve non-violent forms of bystander interventions. In other words, the pattern of defenders provides a promising skillset for constructive forms of bystander intervention. Future research might focus especially on this group in order to better understand the factors that allow people to act as defenders without getting involved in the cycle of violence themselves.

Limitations

The use of a latent profile analysis comes with several limitations. The assumption of homogeneity within profiles may lead to oversimplification of the underlying data (Schmidt et al., 2021). Even though more profiles would not have been conducive to the purpose of the current analysis and did not fit the overall data structure the best, individual patterns among national samples may have been obscured by the generalization of profile sizes. The consistency of profiles across national samples, however, does indicate that there is some degree of validity of the identified profiles across countries. Furthermore, comparisons between included and excluded respondents indicated small to moderate differences for some variables. Excluded participants reported somewhat lower exposure to hate and defending behaviour, but slightly higher acceptance of retaliation. Although these differences were relatively small and the proportion of excluded cases was limited (approximately 10%), this pattern suggests that individuals with lower engagement in online incivility contexts but stronger normative acceptance of retaliation may be slightly underrepresented in the analytic sample. Even though this cross-national approach was successful in validating patterns of experiences and behaviors across countries, the sample is limited to mostly western countries and only includes young adults with men being underrepresented. Collecting and analyzing data that includes more cultural and demographic diversity could contribute to more generalizable findings.

The exclusive focus on young adults limits generalizability but also reflects a demographic group that is highly relevant given the topic, as young adults consistently report the highest levels of internet use (Kumar, 2025) and are especially exposed and vulnerable to incivility (Wang et al., 2019; Politte-Corn et al., 2023). This pattern of frequent internet use is evident in the studied sample and might lead to an underestimation of the share of participants with little or no exposure to online incivility. Younger users also tend to evaluate norm violations less severely than older cohorts (Bormann, 2022), which may affect how they perceive and engage with uncivil discourse.

The implications drawn from the current study may also be limited by the age of the survey data that was used. After 2018, when the data was collected, internet use rose slightly during the COVID-19 pandemic and returned to

similar levels by 2023 (Kumar, 2025), suggesting relatively stable usage patterns. However, major global developments, such as the COVID-19 pandemic, rising political polarization as well as platform-level changes to content moderation, may have influenced the nature of online discourse and how individuals respond to incivility (Wakefield & Wakefield, 2022; Sorrentino et al., 2023). These shifts could affect emotional reactions, engagement behavior, or perceived social norms. Since this study provides insights into user patterns in a pre-pandemic context, its findings should be interpreted with this temporal limitation in mind.

Conclusion

The current research contributes to the existing knowledge about psychological and social predictors that motivate people to participate or intervene in online uncivil discourse. By making use of Latent Profile Analysis, the investigation adds to the understanding of this complex phenomenon by taking a person-centered perspective. The findings underscore the complexity of online incivility and the interconnectedness of various factors influencing individuals' responses to it, while providing new insights as to how perpetration, victimization, bystander intervention, exposure to incivility, and acceptance of retaliation all play a role in shaping individuals' responses to online incivility. As incivility and its negative implications for online communication will continue to affect individual wellbeing and societal discourse, building on current knowledge to deepen the understanding of underlying causes remains highly relevant.

Conflict of Interest

The authors have no conflicts of interest to declare.

Use of AI Services

The authors declare they have used AI services, specifically ChatGPT, for grammar correction and minor style refinements and to edit parts of the analytical script. They carefully reviewed all suggestions from these services to ensure the original meaning, factual accuracy and correctness of the script were preserved.

Data Availability Statement

This study used third party data that the author does not have permission to share.

Authors' Contribution

Emily Ahrens: project administration, conceptualization, data curation, formal analysis, visualization, writing—original draft. **Vladimir Bojarskich**: project administration, conceptualization, methodology, validation, writing—review & editing. **Tobias Rothmund**: supervision, conceptualization, methodology, validation, writing—review & editing.

Acknowledgement

We thank James Hawdon, Pekka Räsänen, and colleagues for the provision of this data. The data collected by Hawdon et al. (2018) was supported by The Institute for Society, Culture, and the Environment at Virginia Tech. The data source had no role in the study design, collection, analysis or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

References

Aiken, M., & Waller, B. D. (2000). Flaming among first-time group support system users. *Information & Management*, *37*(2), 95–100. https://doi.org/10.1016/s0378-7206(99)00036-1

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2013). The "nasty effect:" Online incivility and risk perceptions of emerging technologies. *Journal Of Computer-Mediated Communication*, *19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Babaei, A., Vahedi, S., Imanzadeh, A., & Adib, Y. (2020). Effectiveness of ethical sensitivity skill training on moral disengagement and bullying in adolescent girls. *Social Psychology Research, 10*(37), 123–139. https://doi.org/10.22034/SPR.2020.109700

Bauer, J. (2022). A primer to latent profile and latent class analysis. In Goller, M., Kyndt, E., Paloniemi, S., Damşa, C. (Eds.), *Methods for researching professional learning and development* (pp. 243–268). Springer, Cham. https://doi.org/10.1007/978-3-031-08518-5_11

Bettis, A. H., Burke, T. A., Nesi, J., & Liu, R. T. (2021). Digital technologies for emotion-regulation assessment and intervention: A conceptual review. *Clinical Psychological Science*, *10*(1), 3–26. https://doi.org/10.1177/21677026211011982

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, *41*(S1), 3–33. https://doi.org/10.1111/pops.12670

Bilewicz, M., Soral, W., Świderska, A., Küster, D., Winiewski, M., & Wypych, M. (2025). The emotional drivers of hate speech: Unpacking the central role of contempt in derogatory communication. *Atlantic Journal of Communication*, *33*(5), 766–784. https://doi.org/10.1080/15456870.2025.2525790

Bormann, M. (2022). Perceptions and evaluations of incivility in public online discussions—Insights from focus groups with different online actors. *Frontiers in Political Science*, *4*. https://doi.org/10.3389/fpos.2022.812145

Chen, G. M., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal Of Broadcasting & Electronic Media*, *61*(1), 108–125. https://doi.org/10.1080/08838151.2016.1273922

Cleveland, M. J., Collins, L. M., Lanza, S. T., Greenberg, M. T., & Feinberg, M. E. (2010). Does individual risk moderate the effect of contextual-level protective factors? A latent class analysis of substance use. *Journal of Prevention & Intervention in The Community, 38*(3), 213–228. https://doi.org/10.1080/10852352.2010.486299

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Costello, M., & Hawdon, J. (2018). Who are the online extremists among us? Sociodemographic characteristics, social networking, and online experiences of those who produce online hate materials. *Violence and Gender*, *5*(1), 55–60. https://doi.org/10.1089/vio.2017.0048

DiFranzo, D., Taylor, S. H., Kazerooni, F., Wherry, O. D., & Bazarova, N. N. (2018). Upstanding by design. In *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Article 211). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173785

Festl, R., Scharkow, M., & Quandt, T. (2014). The individual or the group: A multilevel analysis of cyberbullying in school classes. *Human Communication Research*, *41*(4), 535–556. https://doi.org/10.1111/hcre.12056

Frischlich, L., Schatto-Eckrodt, T., Boberg, S., & Wintterlin, F. (2021). Roots of incivility: How personality, media use, and online experiences shape uncivil participation. *Media and Communication*, *9*(1), 195–208. https://doi.org/10.17645/mac.v9i1.3360

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, *12*(2), 167–185. https://doi.org/10.1080/19331681.2014.997416

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852-870). Oxford University Press.

Hansen, T. M., Lindekilde, L., Karg, S. T., Petersen, M. B., & Rasmussen, S. H. R. (2024). Combatting online hate: Crowd moderation and the public goods problem. *Communications*, *49*(3), 444–467. https://doi.org/10.1515/commun-2023-0109

Hawdon, James, John Ryan, Catherine Blaya, Mathew Costello, Vicente Llorent, Atte Oksanen, Ashley Reichelmann, Pekka Räsänen, and Izabela Zych. (2018). "Online Hate in Six Nations: Exposure to and the Production of Online Hate among 18–26-year-olds in Finland, France, Poland, Spain, the United Kingdom and the United States." funded by "The Institute for Society, Culture, and the Environment at Virginia Tech"

Henson, B., Fisher, B. S., & Reyns, B. W. (2019). There is virtually no excuse: The frequency and predictors of college students' bystander intervention behaviors directed at online victimization. *Violence Against Women*, *26*(5), 505–527. https://doi.org/10.1177/1077801219835050

Hood, M., & Duffy, A. (2018). Understanding the relationship between cyber-victimization and cyber-bullying on social network sites: The role of moderating factors. *Personality and Individual Differences*, *133*, 103–108. https://doi.org/10.1016/j.paid.2017.04.004

Hsueh, M., Yogeeswaran, K., & Malinen, S. (2015). "Leave your comment below": Can biased online comments influence our own prejudicial attitudes and behaviors? *Human Communication Research*, *41*(4), 557–576. https://doi.org/10.1111/hcre.12059

Izquierdo-Montero, A., Laforgue-Bullido, N., & Abril-Hervás, D. (2022). Hate speech: A systematic review of scientific production and educational considerations. *Revista Fuentes*, *2*(24), 222–233. https://doi.org/10.12795/revistafuentes.2022.20240

Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the effectiveness of deplatforming as a moderation strategy on twitter. *Proceedings of the ACM On Human-Computer Interaction*, *5*(CSCW2), Article 381. https://doi.org/10.1145/3479525

Kácha, O., Vintr, J., & Brick, C. (2022). Four Europes: Climate change beliefs and attitudes predict behavior and policy preferences using a latent class analysis on 23 countries. *Journal of Environmental Psychology*, *81*, Article 101815. https://doi.org/10.1016/j.jenvp.2022.101815

Krettenauer, T., & Casey, V. (2015). Moral identity development and positive moral emotions: Differences involving authentic and hubristic pride. *Identity*, *15*(3), 173–187. https://doi.org/10.1080/15283488.2015.1023441

Kumar, N. (2025, September 16). *Average screen time statistics 2025 (Global data)*. DemandSage. https://www.demandsage.com/screen-time-statistics/

Laconi, S., Kaliszewska-Czeremska, K., Gnisci, A., Sergi, I., Barke, A., Jeromin, F., Groth, J., Gámez-Guadix, M., Özcan, N. K., Demetrovics, Z., Király, O., Siomos, K., Floros, G., & Kuss, D. J. (2018). Cross-cultural study of problematic internet use in nine European countries. *Computers in Human Behavior*, *84*, 430–440. https://doi.org/10.1016/j.chb.2018.03.020

Lewandowska-Tomaszczyk, B. (2017). Incivility and confrontation in online conflict discourses. *Lodz Papers in Pragmatics*, *13*(2). https://doi.org/10.1515/lpp-2017-0017

Lickel, B., Schmader, T., Curtis, M., Scarnier, M., & Ames, D. R. (2005). Vicarious shame and guilt. *Group Processes & Intergroup Relations*, *8*(2), 145–157. https://doi.org/10.1177/1368430205051064

Lickel, B., Steele, R. R., & Schmader, T. (2011). Group-based shame and guilt: Emerging directions in research. *Social and Personality Psychology Compass*, *5*(3), 153–163. https://doi.org/10.1111/j.1751-9004.2010.00340.x

Lück, J., & Nardi, C. (2019). Incivility in user comments on online news articles: Investigating the role of opinion dissonance for the effects of incivility on attitudes, emotions and the willingness to participate. *Studies in Communication and Media*, *8*(3), 311–337. https://doi.org/10.5771/2192-4007-2019-3-311

Machackova, H., Blaya, C., Bedrosova, M., Smahel, D., & Staksrud, E. (2020). *Children's experiences with cyberhate*. EU Kids Online, The London School of Economics and Political Science. https://doi.org/10.21953/lse.zenkg9xw6pua

Madriaza, P., Hassan, G., Brouillette-Alarie, S., Mounchingam, A. N., Durocher-Corfa, L., Borokhovski, E., Pickup, D., & Paillé, S. (2025). Exposure to hate in online and traditional media: A systematic review and meta-analysis of the impact of this exposure on individuals and communities. *Campbell Systematic Reviews: Better Evidence for a Better World*, *21*(1), Article e70018. https://doi.org/10.1002/cl2.70018

Mamakos, M., & Finkel, E. J. (2023). The social media discourse of engaged partisans is toxic even when politics are irrelevant. *PNAS Nexus*, *2*(10), Article pgad325. https://doi.org/10.1093/pnasnexus/pgad325

Marcotte, D., & Starrs, C. J. (2023). A latent profile analysis of at-risk students in a multilevel prevention program: Who chooses to participate and the impact on outcomes. *Mental Health & Prevention, 31*, Article 200295. https://doi.org/10.1016/j.mhp.2023.200295

Masullo, G. M., Lu, S., & Fadnis, D. (2020). Does online incivility cancel out the spiral of silence? A moderated mediation model of willingness to speak out. *New Media & Society*, *23*(11), 3391–3414. https://doi.org/10.1177/1461444820954194

Miranda, G. A., Welbourne, J. L., & Sariol, A. M. (2020). Feeling shame and guilt when observing workplace incivility: Elicitors and behavioral responses. *Human Resource Development Quarterly*, *31*(4), 371–392. https://doi.org/10.1002/hrdq.21395

Moxey, N., & Bussey, K. (2019). Styles of bystander intervention in cyberbullying incidents. *International Journal of Bullying Prevention*, *2*(1), 6–15. https://doi.org/10.1007/s42380-019-00039-1

Müller, K., & Schwarz, C. (2020). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, *19*(4), 2131–2167. https://doi.org/10.1093/jeea/jvaa045

Ng, Y., Song, Y., Kwon, K. H., & Huang, Y. (2019). Toward an integrative model for online incivility research: A review and synthesis of empirical studies on the antecedents and consequences of uncivil discussions online. *Telematics and Informatics*, *47*, Article 101323. https://doi.org/10.1016/j.tele.2019.101323

Nylund-Gibson, K., & Hart, S. R. (2014). Latent class analysis in prevention science. In Z. Sloboda & H. Petras (Eds.), *Defining Prevention Science* (pp. 493–511). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7424-2_21

Obermaier, M. (2022). Youth on standby? Explaining adolescent and young adult bystanders' intervention against online hate speech. *New Media & Society*, *26*(8), 4785–4807. https://doi.org/10.1177/14614448221125417

Obermaier, M., Schmid, U. K., & Rieger, D. (2023). Too civil to care? How online hate speech against different social groups affects bystander intervention. *European Journal of Criminology*, *20*(3), 817–833. https://doi.org/10.1177/14773708231156328

Obermaier, M., & Schmuck, D. (2022). Youths as targets: Factors of online hate speech victimization among adolescents and young adults. *Journal of Computer-Mediated Communication*, *27*(4), Article zmac012. https://doi.org/10.1093/jcmc/zmac012

Osborne, M. R., Abdurahman, S., Omrani, A., Trager, J., & Dehghani, M. (2025). *Two wrongs is what makes it more right: How retaliatory incivility receives social leniency*. PsyArXiv. https://doi.org/10.31234/osf.io/6fd7h_v5

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society, 6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Politte-Corn, M., Nick, E. A. and Kujawa, A. (2023). Age-related differences in social media use, online social support, and depressive symptoms in adolescents and emerging adults. *Child and Adolescent Mental Health*, *28*(4), 497–503. https://doi.org/10.1111/camh.12640

Postmes, T., Gordijn, E. H., van Zomeren, M., Greijdanus, H., de Vos, B., Täuber, S., & Kamans, E. (2014). Escalation and de-escalation of intergroup conflict: The role of communication within and between groups. In C. K. W. De Dreu (Ed.), *Social conflict within and between groups* (pp. 93–116). Psychology Press.

Rasmussen, S. H. R., Bor, A., & Petersen, M. B. (2024). The offline roots of online hostility: Adult and childhood administrative records correlate with individual-level hostility on Twitter. *Proceedings of the National Academy of Sciences*, *121*(44), Article e2412277121. https://doi.org/10.1073/pnas.2412277121

Rasmussen, S. H. R., & Petersen, M. B. (2023). The event-driven nature of online political hostility: How offline political events make online interactions more hostile. *PNAS Nexus*, *2*(11), Article pgad382. https://doi.org/10.1093/pnasnexus/pgad382

Rhead, R., Elliot, M., & Upham, P. (2018). Using latent class analysis to produce a typology of environmental concern in the UK. *Social Science Research*, *74*, 210–222. https://doi.org/10.1016/j.ssresearch.2018.06.001

Rosenfeld, M. (2003). Hate speech in constitutional jurisprudence: A comparative analysis. *Cardozo Law Review, 24*(4), Article 5. https://larc.cardozo.yu.edu/clr/vol24/iss4/5/

Rothmund, T., & Baumert, A. (2013). Shame on me. *Social Psychological and Personality Science*, *5*(2), 195–202. https://doi.org/10.1177/1948550613488950

Rothmund, T., Farkhari, F., Ziemer, C., & Azevedo, F. (2022). Psychological underpinnings of pandemic denial - patterns of disagreement with scientific experts in the German public during the COVID-19 pandemic. *Public Understanding of Science*, *31*(4), 437–457. https://doi.org/10.1177/09636625211068131

Rösner, L., Winter, S. & Krämer, N. C. (2016). Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior*, *58*, 461–470. https://doi.org/10.1016/j.chb.2016.01.022

Saleem, H. M., & Ruths, D. (2018). The aftermath of disbanding an online hateful community. *arXiv.* https://doi.org/10.48550/arxiv.1804.07354

Salgado, S., De Zúñiga, H. G., Da Silva, P. A., Biscaia, A., Coimbra, M. E., Martins, B., & Francisco, A. P. (2023). Assessing the prevalence and predictors of incivility in online news comments across six countries. *Journalism Practice*, *18*(2), 224–241. https://doi.org/10.1080/17512786.2023.2246431

Schmidt, M. N., Seddig, D., Davidov, E., Mørup, M., Albers, K. J., Bauer, J. M., & Glückstad, F. K. (2021). Latent profile analysis of human values: What is the optimal number of clusters? *Methodology*, *17*(2), 127–148. https://doi.org/10.5964/meth.5479

Scrucca, L., Fraley, C., Murphy, B., & E, R. A. (2023). *Model-based clustering, classification, and density estimation using mclust in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003277965

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Simmons, D. (2022, 13. January). *17 countries with GDPR-like data privacy laws*. Comforte. https://insights.comforte.com/countries-with-gdpr-like-data-privacy-laws

Šincek, D. (2021). The revised version of the Committing and Experiencing Cyber-Violence Scale and its relation to psychosocial functioning and online behavioral problems. *Societies*, *11*(3), Article 107. https://doi.org/10.3390/soc11030107

Sorrentino, A., Sulla, F., Santamato, M., Di Furia, M., Toto, G., & Monacis, L. (2023). Has the COVID-19 pandemic affected cyberbullying and cybervictimization prevalence among children and adolescents? A systematic review. *International Journal of Environmental Research and Public Health*, *20*(10), Article 5825. https://doi.org/10.3390/ijerph20105825

Sullivan, G. B., & Day, C. R. (2019). Collective emotions in celebratory, competitive, and conflictual contexts: Exploring the dynamic relations between group-based and collective pride and shame. *Emotions History Culture Society*, *3*(2), 202–222. https://doi.org/10.1163/2208522x-02010057

Sun, Q., Wojcieszak, M., & Davidson, S. (2021). Over-time trends in incivility on social media: Evidence from political, non-political, and mixed sub-reddits over eleven years. *Frontiers in Political Science*, *3,* Article 741605. https://doi.org/10.3389/fpos.2021.741605

Syrjämäki, A. H., Ilves, M., Kiskola, J., Rantasila, A., Isokoski, P., Olsson, T., & Surakka, V. (2022). Facilitating implicit emotion regulation in online news commenting—An experimental vignette study. *Interacting With Computers*, *34*(5), 129–136. https://doi.org/10.1093/iwc/iwad010

Tangney, J. P., & Fischer, K. W. (Eds.). (1995). *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride*. Guilford Press.

Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The dynamics of political incivility on Twitter. *SAGE Open*, *10*(2). https://doi.org/10.1177/2158244020919447

Thompsen, P. A., & Foulger, D. A. (1996). Effects of pictographs and quoting on flaming in electronic mail. *Computers in Human Behavior*, *12*(2), 225–243. https://doi.org/10.1016/0747-5632(96)00004-0

Tracy, J. L., & Robins, R. W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry*, *15*(2), 103–125. https://doi.org/10.1207/s15327965pli1502_01

Tracy, J. L., & Robins, R. W. (2006). Appraisal antecedents of shame and guilt: Support for a theoretical model. *Personality and Social Psychology Bulletin*, *32*(10), 1339–1351. https://doi.org/10.1177/0146167206290212

Van Zomeren, M., D'Amore, C., Pauls, I. L., Shuman, E., & Leal, A. (2023). The intergroup value protection model: A theoretically integrative and dynamic approach to intergroup conflict escalation in democratic societies. *Personality and Social Psychology Review*, *28*(2), 225–248. https://doi.org/10.1177/10888683231192120

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*(4), 223–229. https://doi.org/10.1089/cyber.2021.0159

Wachs, S., Mazzone, A., Milosevic, T., Wright, M. F., Blaya, C., Gámez-Guadix, M., & Norman, J. O. (2021). Online correlates of cyberhate involvement among young people from ten European countries: An application of the routine activity and problem behavior theory. *Computers in Human Behavior*, *123*, Article 106872. https://doi.org/10.1016/j.chb.2021.106872

Wachs, S., Wettstein, A., Bilz, L., Espelage, D. L., Wright, M. F., & Gámez-Guadix, M. (2024). Individual and contextual correlates of latent bystander profiles toward racist hate speech: A multilevel person-centered approach. *Journal of Youth and Adolescence*, *53*(6), 1271–1286. https://doi.org/10.1007/s10964-024-01968-x

Wachs, S., Wright, M. F., Sittichai, R., Singh, R., Biswal, R., Kim, E., Yang, S., Gámez-Guadix, M., Almendros, C., Flora, K., Daskalou, V., & Maziridou, E. (2019). Associations between witnessing and perpetrating online hate in eight countries: The buffering effects of problem-focused coping. *International Journal of Environmental Research and Public Health*, *16*(20), Article 3992. https://doi.org/10.3390/ijerph16203992

Wakefield, R. L., & Wakefield, K. (2022). The antecedents and consequences of intergroup affective polarisation on social media. *Information Systems Journal*, *33*(3), 640–668. https://doi.org/10.1111/isj.12419

Wang, M.-J., Yogeeswaran, K., Andrews, N. P., Hawi, D. R., & Sibley, C. G. (2019). How common is cyberbullying among adults? Exploring gender, ethnic, and age differences in the prevalence of cyberbullying. *Cyberpsychology, Behavior, and Social Networking*, *22*(11), 736-741. https://doi.org/10.1089/cyber.2019.0146

Warner, W., & Hirschberg, J. (2012). Detecting hate speech on the world wide web. In *Proceedings of the Second Workshop on Language in Social Media* (pp. 19–26). Association for Computational Linguistics. https://aclanthology.org/W12-2103/

Windisch, S., Wiedlitzka, S., Olaghere, A., & Jenaway, E. (2022). Online interventions for reducing hate speech and cyberhate: A systematic review. *Campbell Systematic Reviews*, *18*(2), Article e1243. https://doi.org/10.1002/cl2.1243

World Bank Group. (2024). *World bank open data*. https://data.worldbank.org/indicator/IT.NET.USER.ZS

About Authors

Emily Ahrens is a PhD candidate in Political Psychology at the Institute of Communication Science and a researcher at the Centre for Right-Wing Extremism Research, Democracy Education and Social Integration (KomRex) at the Friedrich Schiller University Jena in Germany. Her research is centered around the psychological underpinnings of support for democracy and accompanying phenomena, such as hate speech.

https://orcid.org/0009-0004-4899-120X

Vladimir Bojarskich is a PhD candidate in Psychology of Communication and Media Use at the Institute of Communication Science and a researcher at the Centre for Right-Wing Extremism Research, Democracy Education and Social Integration (KomRex) at the Friedrich Schiller University Jena in Germany. In his research, he is interested in understanding the social, moral, and ideological motives that drive intergroup conflict (e.g., hate speech) and people's (dis-)engagement with contemporary political issues (e.g., climate change).

https://orcid.org/0009-0005-7477-9306

Tobias Rothmund is Professor for Psychology of Communication and Media Use and Director of the Centre for Right-Wing Extremism Research, Democracy Education and Social Integration (KomRex) at the Friedrich Schiller University in Jena, Germany. His research interests are located in the intersection between personality research, political psychology and mass media communication. Current work addresses topics such as right-wing populism, political ideology and justice sensitivity, justice perceptions in mass media communication, motivated science reception and media violence.

https://orcid.org/0000-0003-2979-5129

Correspondence to
Emily Ahrens, Institute for Communication Science, Ernst-Abbe-Platz 8, 07743 Jena, Germany, emily.ahrens@uni-jena.de