

Jiménez-Díaz, O., Mora-Merchán, J. A., Elipe, P., & Del Rey, R. (2025). Understanding bystanders' behaviour to cyberhate: An explanatory approach from the theory of normative social behaviour. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 19(5), Article 10. <https://doi.org/10.5817/CP2025-5-10>

Understanding Bystanders' Behaviour to Cyberhate: An Explanatory Approach From the Theory of Normative Social Behaviour

Olga Jiménez-Díaz¹, Joaquín A. Mora-Merchán¹, Paz Elipe², & Rosario Del Rey¹

¹ Department of Developmental and Educational Psychology, Universidad de Sevilla, Seville, Spain

² Department of Psychology, Faculty of Humanities and Educational Sciences, Universidad de Jaén, Jaén, Spain

Abstract

Cyberhate is a new form of cyberviolence among adolescents with high social and scientific concern. Although previous research has examined aggressors and victims, far less attention has been given to bystanders. Examining how adolescents respond when witnessing cyberhate, and the factors that shape these behaviours, is essential within a solid theoretical framework. For that, this study aims to 1) find out the responses of adolescent bystanders to cyberhate; 2) explore whether the Theory of Normative Social Behaviour (TNSB) can be useful as explanatory framework of adolescents' responses; and 3) examine whether the TNSB shows a better explanatory capacity with the inclusion of behavioural (i.e., toxic online disinhibition), personal (i.e., empathy) and contextual variables (i.e., online parental supervision) compared to the model based solely on social norms. The study included 2,539 Spanish students (49.1% boys, 49.2% girls, 1.7% other) aged 11–18 years ($M = 14.07$; $SD = 1.39$). Structural equation models were applied to test the TNSB. Findings showed adolescents predominantly use defending behaviour (supporting victims), followed by passive (not getting involved) and reinforcing behaviour (supporting aggressors). Good fits and explanatory variances were found for the three behaviours, especially when including toxic online disinhibition, empathy, and online parental supervision. Results highlight the importance of subjective norms and empathy in all bystanders' behaviours. In reinforcing and passive behaviours, friends' and family's injunctive norms, descriptive norms, and toxic online disinhibition were also relevant. Online parental supervision was also related to defending and passive behaviours. Likewise, in passive behaviour, the collective norms were significant. These findings support the explanatory validity of the TNSB on cyberhate bystanders' behaviour, especially for active behaviours, providing relevant results for future psychoeducational programs.

Keywords: cyberhate; bystanders; adolescents; theory of normative social behaviour

Editorial Record

First submission received:
October 14, 2024

Revisions received:
April 28, 2024
September 2, 2025

Accepted for publication:
September 3, 2025

Editor in charge:
Joris Van Ouytsel

Introduction

Hate is an intense aversion towards something or someone that is often accompanied by feelings such as resentment, anger and disgust, and involves a negative evaluation (Staub, 2005). Hate does not always manifest itself overtly, but when it does, it may take the form of physical violence, denigration or verbal attacks (Bauman et al., 2021). As has occurred with other forms of violence in physical environments, the increased use of social

media has led these intentional and violent communicative expressions of hate to migrate to virtual space, transforming into a new form of cyberviolence known as cyberhate (Bauman et al., 2021; Costello et al., 2023).

Cyberhate, also known as online or digital hate, refers to any form of hate expression shared through digital platforms (e.g., text, images, or audiovisual content such as memes and hoaxes) that harms, humiliates, or incites hatred and hostility toward an individual or group based on prejudice or stereotypes (Bührer et al., 2024; Jiménez-Díaz & Del Rey, 2025). Research suggests that adolescents are mostly involved as bystanders in this kind of cyberviolence (Wachs, Wettstein, et al., 2024). Cyberhate's bystanders—or those who are exposed but not necessarily targeted by it (Bedrosova et al., 2023)—represent approximately 74% to 84% of the adolescent population (Hawdon et al., 2023; Obermaier, 2024). Although not directly targeted, any involvement in the phenomenon leads to negative consequences in various aspects of an individual's life (Llorent et al., 2023) and increases prejudice (March & Marrington, 2019). Therefore, it is essential that cyberhate's study not only prioritises, as it has been doing to date, studying victims (e.g., Costello et al., 2017; Wachs et al., 2021) and perpetrators (e.g., Blaya & Audrin, 2019; Hawdon et al., 2023), but also bystanders and their role in the phenomenon.

The role of bystanders is particularly complex, encompassing a wide spectrum of potential behaviours, from passive inaction to active reinforcement of the perpetrator (Wachs, Bilz, et al., 2024). Traditionally, bystanders have been classified into four profiles depending on their behaviour when bystanding a violent situation: defenders, who defend the victims or help to stop the situation; passive, who ignore it; reinforcers, who encourage the aggressor by showing approval; and assistants, who directly assist and join the aggressor (Salmivalli et al., 1996). Assistants and reinforcers are commonly grouped under the same profile, as both adopt roles that support the aggressor through reinforcing behaviours (Álvarez-García et al., 2021; Thornberg & Jungert, 2013). Previous research has analysed the bystanders' profiles of offline racist hate speech, obtaining similar results to the traditional literature, indicating that there are adolescents who are mainly defenders (47.3%), followed by passives (34.2%), and revengers (9.8%) or contributors (8.6%), who are those who reinforce the action in a negative way (Wachs, Wettstein, et al., 2024). In its online counterpart, it seems to follow the same line. Although research analysing bystander behaviours indicates that between 30% to 42% of youths never intervene in these situations (Costello et al., 2023; Obermaier, 2024), when they do intervene, it is mostly to support the victim or report the offensive behaviour (Obermaier, 2024). However, there is still a lack of evidence on how adolescents, as bystanders, may intervene negatively by reinforcing aggression in online hate contexts, highlighting the need for further research in this area. Based on the reviewed empirical research, this study hypothesises that bystanders will primarily engage in defending behaviour, followed by passive and reinforcing behaviour when witnessing cyberhate (H1).

The type of bystanders' response directly impacts the persistence and spread of cyberhate (Garland et al., 2022; Obermaier et al., 2023). Their behaviour is crucial, as cyberhate—much like cyberbullying—is not merely a dyadic aggressor-victim dynamic, but rather a psychosocial phenomenon in which bystanders' reactions can significantly influence its escalation or mitigation (Menesini & Salmivalli, 2017). Then, it is essential to understand how they react and the factors that influence their decision-making in relation to one behaviour or another. Previous research has identified several factors that influence bystander behaviour against cyberhate. For instance, factors such as strong online collective efficacy have been shown to increase the likelihood of intervention (Costello et al., 2023); high exposure to online hate tends to discourage bystanders' willingness to act (Obermaier, 2024), and classrooms environments characterized by frequent hate speech may even encourage bystanders to reinforce aggressive behaviours (Wachs, Wettstein, et al., 2024). While these insights provide a foundational framework, they may not fully capture the complex dynamics of bystander behaviour against cyberhate, possibly because these variables have been studied in a casual or isolated manner. A comprehensive theoretical model is needed to integrate existing knowledge, elucidate the psychological mechanisms driving intervention and inaction, and examine how these variables interact to shape bystander behaviour.

In this line, the Theory of Social Normative Behaviour (TNSB) emerges as a suitable framework for understanding the behaviour of bystanders when they witness cyberhate. This theory postulates that social norms influence people who tend to behave according to them to avoid social rejection and gain approval within their group (Rimal & Real, 2005; Rimal & Yilma, 2022). Prior research into other forms of cyberviolence, such as cyberbullying, has revealed the essential role played by social norms within close environments in shaping bystander behaviour (e.g., Bastiaensens et al., 2016; Domínguez-Hernández et al., 2018). Therefore, it could be assumed that social norms could also play an important role in cyberhate.

According to Rimal and Yilma (2022), social norms comprise subjective, injunctive, descriptive, and collective norms. Subjective norms refer to individuals' perceptions of what others expect them to do, while injunctive norms involve perceived social approval or disapproval of certain behaviours (Chung & Rimal, 2016). Research has demonstrated the influence of these norms on cyberhate-related behaviours. For instance, injunctive anti-hate speech norms are negatively associated with perpetration and mitigate the link between witnessing and engaging in hate speech (Wachs, Wettstein, et al., 2022). Especially concerning bystander behaviour, the perception of cyberhate norms, when reinforced by a supportive environment that discourages harmful behaviours, has been found to encourage defending intervention in cyberhate situations (Wachs, Castellanos, et al., 2023). Thus, it is reasonable to expect that adolescents who perceive their peers or significant others as disapproving of aggression and valuing intervention are more likely to take action. Conversely, when they perceive tolerance or even endorsement of cyberhate, combined with a lack of social support discouraging such behaviours, they may be less inclined to intervene or may even passively accept or reinforce the aggression.

On the other hand, descriptive norms refer to perceptions about the prevalence of behaviour, whereas collective norms refer to the actual prevalence of the behaviour (Chung & Rimal, 2016). Adolescents' perceptions of how common hateful behaviour on the Internet is among peers have been previously associated with cyberhate exposure, highlighting the role of perceived prevalence in shaping online hate behaviours (Turner et al., 2023). Evidence suggests that these norms are also closely related to bystander behaviour in the context of cyberhate. The frequency with which individuals engage in a behaviour serves as an implicit yet reliable indicator of its perceived prevalence in their social environment, signalling its normalisation and, therefore, increasing the likelihood of performing it (Van Ouytsel et al., 2017). Research indicates that adolescents with prior experience as bystanders or victims of the phenomenon are less likely to intervene (Obermaier, 2024; Rudnicki et al., 2023). Conversely, classroom environments where hate speech is frequent may even encourage bystanders to reinforce aggressive behaviours (Wachs, Wettstein, et al., 2024). These findings suggest that individual perceptions of behavioural frequency and its prevalence within immediate social contexts shape bystander behaviours by fostering the normalisation of cyberhate. Therefore, it is hypothesised that social norms (i.e., subjective, injunctive, descriptive, and collective norms) reflecting a higher acceptance of cyberviolence will be directly associated with passive or reinforcing bystander behaviour, whereas norms with lower acceptance will be linked to defending behaviours (H2).

According to Rimal and Yilma (2022), behaviour is not only influenced by social norms but also by behavioural attributes, personal characteristics, and contextual factors, which have been incorporated to expand the TNSB. These factors, alongside social norms, may provide a deeper understanding of bystander behaviour in cyberhate contexts. A previous systematic review on bystander behaviour suggests that multiple individual and contextual factors play a role in shaping behaviours related to online hate (Rudnicki et al., 2023). In this regard, it is hypothesised that incorporating behavioural, personal, and contextual variables will significantly enhance the explanatory power of cyberhate bystander behaviour models and influence the relationship between social norms and bystander behaviours and intentions (H3).

Based on Rimal and Yilma (2022), certain inherent behaviour characteristics can increase or decrease the effectiveness of normative influence. Given that cyberhate is an inherently online phenomenon occurring in a context marked by anonymity, invisibility, asynchrony, a false sense of impunity, and a lack of eye contact (Suler, 2004), toxic online disinhibition may be a key attribute. Toxic online disinhibition refers to reducing social constraints and inhibitions usually experienced in face-to-face interactions. This causes individuals to perceive themselves as part of an anonymous mass and, therefore, not feel responsible for their actions (Suler, 2004). As a behavioural attribute, toxic online disinhibition describes the tendency of individuals to behave in a disinhibited and aggressive manner in digital environments, which is a core characteristic of the phenomenon under study. Previous research shows the importance of toxic online disinhibition in the relation between being a cyberhate bystander and becoming a perpetrator, underscoring its importance as a key factor in understanding bystander behaviour in the phenomenon (Jiménez-Díaz et al., 2024; Wachs & Wright, 2018). Therefore, it could be assumed that this attribute may lead adolescents to inhibit the need to adapt to social norms that influence them in face-to-face environments and not feel responsible for intervening when they witness episodes of cyberviolence, thus becoming passive or even more easily provoking a reinforcing behaviour (H3a).

Another key factor, in this case personal, is empathy. There is evidence that empathy is inversely related to the perpetration and acceptance of online hate speech (Celuch et al., 2022; Wachs, Bilz, et al., 2022), suggesting that high levels of empathy may act as a protective factor when social norms are inappropriate, inhibiting the

normalisation of cyberhate and promoting defending behaviours. This is evidenced by previous research findings that have shown empathy to be positively correlated with willingness to help as a bystander of cyberhate (Costello et al., 2023; Rudnicki et al., 2023) and hate speech (Wachs, Krause, et al., 2023). However, it is worth noting that this research has measured empathy in face-to-face settings without specifically considering its application in digital environments, where contextual characteristics may influence and limit it (Voggeser et al., 2018). Therefore, it is essential to delve into the influence of online empathy to understand bystanders' behaviours to cyberhate. In this sense, it is hypothesised that higher levels of online empathy facilitate defending behaviour, whereas lower levels facilitate reinforcing and passive behaviour (H3b).

Finally, a relevant contextual factor in this dynamic is online parental supervision. Online parental supervision is understood as the activities of parents aimed at protecting their children from exposure to risky activities and dangers online (Livingstone et al., 2015), so it is to be expected that adolescents' behaviour in virtual environments is influenced by it. Parental supervision may be directly related to how adolescents bystanders respond to episodes of cyberhate, possibly by increasing the use of effective coping strategies (Wright et al., 2021). However, it could also act indirectly by decreasing the influence of social norms on relevant others, such as peers, when inappropriate. Nevertheless, this association has not been explored to date. As such, parental supervision emerges as a potentially important factor that, along with empathy and toxic online disinhibition, may influence social norms and help explain the behaviours of adolescent bystanders to cyberhate. Then, it is hypothesised that stronger perceived parental supervision facilitates defending behaviour, while weaker perceptions facilitate reinforcing and passive behaviour (H3c).

Current Study

Cyberhate bystanders play a crucial role in its dynamics. Although significant progress has been made in understanding their behaviour, further research is needed to delve deeper into the factors that shape it. In particular, this should be done within a consolidated theoretical framework that allows for a comprehensive understanding of how these variables interact to explain bystander behaviour. Therefore, the present study was carried out with the objectives of 1) finding out the behaviours of adolescent bystanders to cyberhate; 2) exploring whether the TNSB can be useful as explanatory framework of adolescents' behaviours; and 3) examining whether the TNSB shows a better explanatory capacity with the inclusion of behavioural (i.e., toxic online disinhibition), personal (i.e., empathy) and contextual variables (i.e., online parental supervision).

Based on the scientific literature, the following research hypotheses are proposed:

H1: Bystanders will engage primarily in defending behaviour, followed by passive and reinforcing behaviour when seeing cyberhate.

H2: Social norms reflecting higher acceptance of cyberviolence will be directly associated with passive or reinforcing bystander behaviour (H2a), whereas norms with lower acceptance will be linked to defending behaviours (H2b).

H3: Including behavioural (i.e., toxic online disinhibition), personal (i.e., online empathy), and contextual variables (i.e., online parental supervision) will significantly enhance the explanatory power of cyberhate bystander behaviour models and influence the relationship between social norms and bystander behaviour and intentions. Especially, (H3a) higher levels of toxic online disinhibition decrease the likelihood of defending behaviour, while increasing passive and reinforcing behaviour; (H3b) higher levels of online empathy facilitate defending behaviour, whereas lower levels facilitate reinforcing and passive behaviour; (H3c) stronger perceived parental supervision facilitates defending behaviour, while weaker perceptions facilitate reinforcing and passive behaviour.

Methods

Participants

This cross-sectional study, conducted during the first half of 2023, involved 2,539 students (49.1% boys, 49.2% girls, 1.7% others) aged 11–18 years ($M = 14.07$, $SD = 1.39$) from 18 middle and high schools in three different provinces of Andalusia, Spain. Of the participating schools, nine were public and nine were publicly funded private schools. According to publicly available information on the participating schools, it has been estimated that of the eighteen schools included in the study, sixteen are located in middle-class areas, and two are in deprived areas. The sample was selected through convenience sampling, with schools being invited to participate voluntarily.

Instruments

The Student Bystander Behaviour Scale

The Student Bystander Behaviour Scale (*SBBS*; Thornberg & Jungert, 2013), adapted by Álvarez-García et al. (2021), measured the bystander behaviour adapted to cyberhate. This instrument comprises 10 Likert-style items with five response options (0 = *never* to 4 = *daily*) that ask participants how they responded to, or think they would have responded to, episodes of cyberhate in the current academic year. Four of the items referred to the defending behaviour (e.g., *I talked to the aggressor at another time to get him/her to stop*; e.g., *through their wall, chat or private messages*), three of them to the passive behaviour (e.g., *I avoided getting involved, I don't support either one or the other*) and three to the reinforcer behaviour (e.g., *I joined the message or comment thread and started picking on the person myself*). CFA showed a good structure (χ^2 S-B (32) = 310.1209; $p < .001$; RMSEA = .07; 90% CI [.043, .056], SRMR = .06; CFI = .97; NNFI = .95). Reliability was good for the defending behaviour (CR = 0.81; H Coefficient = .81; Cronbach's $\alpha = .81$) and passive behaviour (CR = 0.78; H Coefficient = .77; Cronbach's $\alpha = .76$), and acceptable for the reinforcing behaviour (CR = 0.80; H Coefficient = .70; Cronbach's $\alpha = .70$).

Cyberhate Involvement

First, participants were provided with the definition of the phenomenon (i.e., *Cyberhate is the set of messages or comments that express hatred, denigrate, humiliate or discriminate against a person or group of people through the Internet and social networks*). After that, following scale of cyberhate (Jiménez-Díaz et al., 2025), four direct questions about having seen, sent, received, or having intentionally searched to view cyberhate in the last 12 months were used. These items were Likert-type with five response options ranging from 0 = *never* to 4 = *daily*. To assess its factor structure, a Confirmatory Factor Analysis (CFA) was performed and showed a good unidimensional structure; χ^2 S-B (2) = 36.8442; $p < .001$; RMSEA = .08; 90% CI [.061, .107], SRMR = .04; CFI = .98; NNFI = .95; Construct Reliability (CR) = 0.82; H Coefficient = .79; Cronbach's $\alpha = .79$. The items also represented the descriptive norms, and they were used to calculate the collective norms.

Collective Norms

In line with the recommendations of Rimal and Yilma (2022), collective norms were calculated at the school level using the non-self-mean for each cyberhate behaviour item (i.e., *see, send, receive, search*). For each participant, their individual score on a given item was subtracted from the school-wide mean for that item. This difference was then divided by $n - 1$ (where n is the number of individuals in the school), yielding a measure of the perceived collective norm for each behaviour, excluding self-influence. Reliability was strong across items: Construct Reliability (CR) = 0.88, H Coefficient = .86, and Cronbach's $\alpha = .81$.

Friends' Injunctive Norms

They were measured through five-point Likert-type items (0 = *never* to 4 = *daily*) that included four items from Sasson and Mesch's (2016) scale on subjective norms of peers approving of online risky behaviour (e.g., *Most of*

my friends think it's OK to upload an offensive video) and a final item from Bastiaensens et al. (2016) to measure the injunctive norm of friends approving of cyberbullying, in this study adapted to cyberhate (i.e., *Most of my friends think it is OK to act violently towards someone online or on a mobile phone*). The CFA showed a good unidimensional structure: CFA; χ^2 S-B (5) = 86.6779; $p < .001$; RMSEA = .08; 90% CI [.065, .094], SRMR = .03; CFI = .98; NNFI = .97; CR = 0.84; H coefficient = .84; Cronbach's α = .83.

Family's Injunctive Norms

They were measured through five-point Likert-type items (0 = *never* to 4 = *daily*) that included four items from Sasson and Mesch's (2016) scale on parental subjective norms of online risk behaviour disapproval (e.g., *My family thinks it is not worth sending an offensive message to someone online*) and a final item from Bastiaensens et al. (2016) to measure the injunctive norm of cyberbullying disapproval, in this study adapted to cyberhate (i.e., *My family would not approve of me acting violently online*). The CFA showed a good unidimensional structure: CFA; χ^2 S-B (5) = 117.3508; $p < .001$; RMSEA = .08; 90% CI [.078, .107], SRMR = .04; CFI = .98; NNFI = .97; CR = 0.90; H Coefficient = .89; Cronbach's α = .89.

Teachers' Injunctive Norms

The previous scale, referring to parents' subjective norms towards online risky behaviour disapproval (e.g., *My teachers think it is not OK to act violently online*), was adapted to the teachers' context. The CFA showed a good unidimensional structure: CFA; χ^2 S-B (5) = 38.3097; $p < .001$; RMSEA = .07; SRMR = .05; 90% CI [.036, .066], CFI = .98; NNFI = .97; CR = 0.84; H coefficient = .84; Cronbach's α = .84.

Subjective Norms

They were measured with a scale of three seven-choice Likert-type items (0 = *strongly disagree* to 6 = *strongly agree*) proposed by Hayashi and Tahmasbi (2022; e.g., *Those people who are important to me would want me to help someone who suffers violence on the Internet*). Reliability was good (CR = 0.81; H coefficient = .81; Cronbach's α = .81).

Toxic Online Disinhibition

Four items from the Online Disinhibition Scale (Udris, 2014) were used to measure toxic online disinhibition (e.g., *I don't mind writing insults online since my identity is anonymous*). The items were Likert-type with four response options (0 = *disagree* to 3 = *agree*). The CFA showed a correct unidimensional structure: χ^2 S-B (2) = 156.9170; $p < .001$; RMSEA = .07; SRMR = .07; 90% CI [.053, .099], CFI = .94; NNFI = .91; CR = 0.79; H Coefficient = .78; Cronbach's α = .78.

Empathy

Empathy was measured with five items from Hayashi and Tahmasbi's (2022) scale concerning empathy in virtual contexts (e.g., *I usually feel bad when I see violence on the Internet*). The items were seven-choice Likert-type (0 = *strongly disagree* to 6 = *strongly agree*). The CFA showed good results: χ^2 S-B (2) = 52.7053; $p < .001$; RMSEA = .06; 90% CI [.048, .079], SRMR = .02; CFI = .99; NNFI = .98; CR = .92; H Coefficient = .91; Cronbach's α = .91.

Online Parental Supervision

The Martin-Criado et al. (2021) scale of parental monitoring of Internet use was used, consisting of four 5-choice Likert-type items (0 = *never* to 4 = *always*; e.g., *My family monitors my use of new technologies*). The CFA showed a good unidimensional structure: χ^2 S-B (2) = 52.8085; $p < .001$; RMSEA = .07; 90% CI [.049, .097], SRMR = .03; CFI = .99; NNFI = .97; CR = .74; H Coefficient = .78; Cronbach's α = .73.

All instruments used in the study are available at <https://doi.org/10.6084/m9.figshare.30383161.v1>.

Procedure

This research has been carried out by the ethical standards of the APA and has been approved by the Andalusian Biomedical Research Ethics Coordinating Committee (2563-N-20).

Firstly, the school management teams were contacted to propose their participation in the project and to provide them with information about it. The study's objectives and the instruments used were presented and explained to the head teachers and the school council of the corresponding school, who positively evaluated the proposal. Once the approval of the schools had been obtained, the consent of the families was requested and processed by the schools. Subsequently, data collection began in January 2023 and ended in June 2023. Authorised students received instructions on how to complete the survey on paper and pencil. They were informed of the voluntary, anonymous, and confidential nature of the research, their right to withdraw at any time and the importance of answering all questions honestly. The survey took approximately 60 minutes and was supervised by the teaching staff and the research team.

Data Analysis

First, after coding and recoding the data in the statistical software SPSS 29, the data were exported to EQS v6.4. and the psychometric properties of the scales were examined with confirmatory factor analysis (CFA), considering the proposed fit indices for categorical variables (Hu & Bentler, 1999).

To address the first objective, means, standard deviations, and correlations among the different types of bystander behaviours were calculated to explore how adolescent bystanders respond to cyberhate. In addition, frequency descriptives were generated for the four types of cyberhate behaviours, along with means and standard deviations for the variables included in the models (i.e., social norms, toxic online disinhibition, empathy, and online parental supervision).

In line with the second objective, aiming to assess the explanatory capacity of the TNSB, a structural equation model on the direct relation between social norms and the three types of bystanders' behaviours (defending, passive, and reinforcing) was estimated. In the present study, social norms comprised six latent variables: subjective norms, friends, family, and teacher injunctive norms, descriptive norms, and collective norms. The variables referring to subjective, friends, family, and teacher injunctive norms were calculated with the mean of the items composing each of them. The latent variable of descriptive norms was composed of the four items of cyberhate involvement (seeing, sending, receiving, and searching), while the latent variable of collective norms was made up of the four collective norm variables corresponding to each cyberhate behaviour. Bystanders' behaviour types (defending, passive, and reinforcing) were also included in the model as latent variables.

For the third objective of the study, a second model was conducted for each type of bystanders' behaviour, where toxic online disinhibition (behavioural attribute), empathy (personal characteristics) and parental supervision (contextual variables) were incorporated as additional variables to examine whether the TNSB provides greater explanatory capacity (see Figure 1). Those variables were also included in the model as latent variables. These models were estimated using the Least Likelihood Robust method, appropriate to the categorical nature of the variables under study. The fit of the models has been tested with the following indices: Satorra-Bentler scaled chi-square (χ^2 S-B; Satorra & Bentler, 2001); the comparative fit index (CFI) and the non-normality fit index (NNFI; ≥ 0.90 is adequate; ≥ 0.95 is optimal); the root mean square error of approximation (RMSEA) and the root mean square residual (SRMR; ≤ 0.08 is adequate; ≤ 0.05 is optimal; Hu & Bentler, 1999). Also, relations with $p < .05$ are considered significant.

[illegible]

Firstly, adolescents were mainly involved in cyberhate by viewing hate messages (80.8%), receiving them (37.9%), sending them (24.4%), and, to a lesser extent, intentionally searching to view cyberhate (22.3%). Regarding bystanders' behaviours to cyberhate episodes, participants reported the highest mean for defending behaviour ($M = 2.00$, $SD = 1.19$), followed by passive ($M = 1.44$, $SD = 1.15$) and reinforcing behaviours ($M = 0.31$, $SD = 0.63$; see Table 1). Defending behaviour was negatively correlated with both reinforcing behaviour ($r = -.12$, $p < .001$), and passive behaviour ($r = -.27$, $p < .001$). Additionally, passive behaviour was positively correlated with reinforcing behaviour ($r = .22$, $p < .001$). Table 2 presents the descriptive analyses of the variables included in the models.

Bystander Behaviour Item	<i>n</i>	<i>M</i>	<i>SD</i>
Defending Behaviour	2,123	2.00	1.19
I encouraged the victim to tell a trusted adult (e.g., family member or teacher).	2,191	2.40	1.55
I told a trusted adult (e.g., family, older sibling, teacher) or reported it on the platform.	2,193	1.99	1.55
I contacted someone I know to try to stop it.	2,164	1.82	1.43
I talked to the aggressor later to ask them to stop (e.g., via chat or private message).	2,195	1.82	1.42
Passive Behaviour	2,139	1.44	1.15
I avoided getting involved, I don't support either one or the other.	2,191	1.61	1.44
I got away from the situation (e.g., muted or unfollowed content).	2,179	1.42	1.38
I did nothing. I continued doing whatever I was doing, because it's not about me.	2,181	1.33	1.39
Reinforcing Behaviour	2,172	0.31	0.63
I read the rest of the messages because they were fun or entertaining.	2,191	0.53	1.01
I joined the comment thread and also attacked the victim.	2,186	0.24	0.72
I laughed and joined in to encourage the aggressor to continue.	2,196	0.18	0.64

Note. $N = 2,539$. Response scale ranged from 0 (*never*) to 4 (*always*).

Table 2. Descriptives of Social Norms, Toxic Online Disinhibition, Empathy and Online Parental Supervision.

	<i>n</i>	<i>M</i>	<i>SD</i>
Social norms			
Friends' injunctive norms	2,363	0.78	0.78
Teachers' injunctive norms	2,375	3.05	1.02
Family's injunctive norms	2,389	3.28	0.91
Subjective norms	2,226	4.51	1.52
Additional variables			
Toxic online disinhibition	2,418	2.04	1.12
Empathy	2,161	4.34	1.55
Online parental supervision	2,413	2.04	1.12

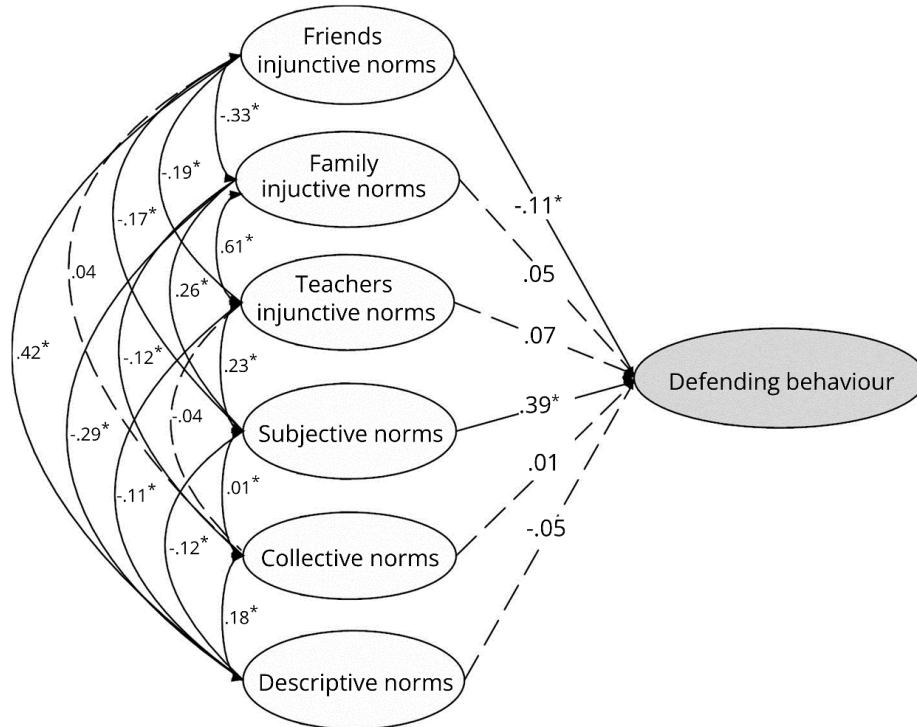
Note. *N* = 2,539. For friends', teachers' and family's injunctive norms, toxic online disinhibition, and online parental supervision: 0 = *totally disagree* and 4 = *totally agree*; for subjective norms and empathy: 0 = *totally disagree* and 6 = *totally agree*.

TSNB Model: Social Norms and Bystanders' Behaviours

Based on the second objective, structural equation models have been estimated with the direct relation between social norms and each type of bystander's behaviour to cyberhate. The fit indices showed a good fit for the defending behaviour: χ^2 S-B (385) = 2173.0042, $p < .001$, RMSEA = .04, 90% CI [.047, .052], SRMR = .04, CFI = .95, NNFI = .95, with an explained variance of 23.3%; for the passive behaviour: χ^2 S-B (357) = 2023.4061, $p < .001$, RMSEA = .05, 90% CI [.053, .058], SRMR = .04, CFI = .94, NNFI = .94, explaining the 8.3%; and for the reinforcer: χ^2 S-B (356) = 2264.9601, $p < .001$, RMSEA = .06, 90% CI [.060, .064], SRMR = .04, CFI = .93, NNFI = .92, explaining the 26.9%.

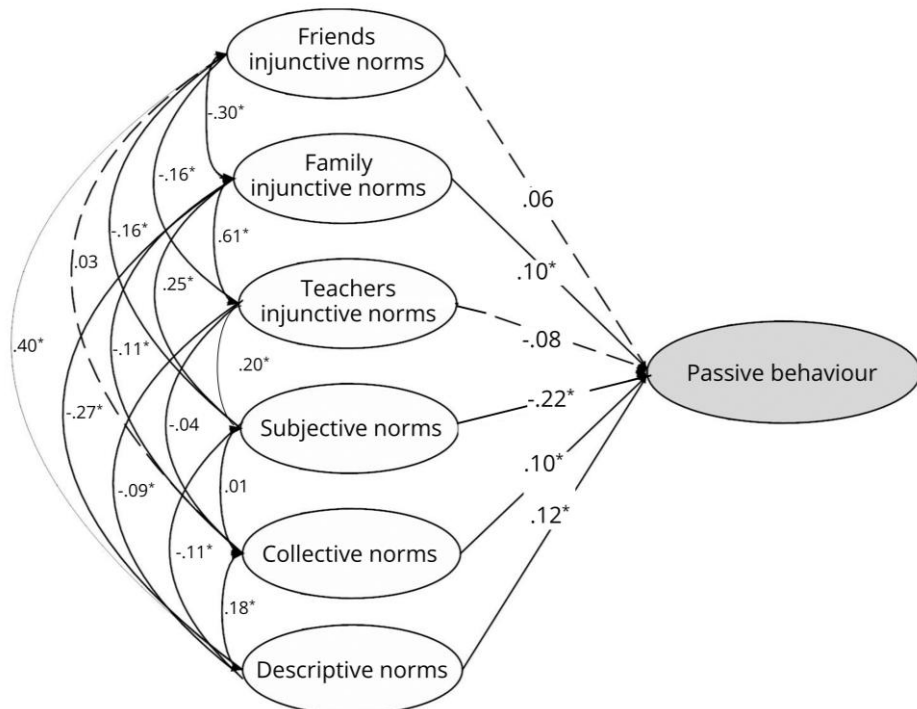
In the defending behaviour, the relations between variables and the β coefficients (see Figure 2) show significant relations with subjective norms positively ($\beta = .39$) and with friends' injunctive norms inversely ($\beta = -.11$). Passive responding was significantly positively related to descriptive norms ($\beta = .12$), family's injunctive norms ($\beta = .10$) and collective norms ($\beta = .10$), and inversely related to subjective norms ($\beta = -.22$; see Figure 3). Finally, reinforcing behaviour was significantly and positively related to descriptive ($\beta = .36$) and friends' injunctive norms ($\beta = .13$), and inversely related to subjective ($\beta = -.17$) and family's injunctive norms ($\beta = -.02$; see Figure 4). The coefficients of the three models are presented in Table 3.

Figure 2. Graphical Representation of the Structural Equation Model (SEM) for Defending Behaviour.



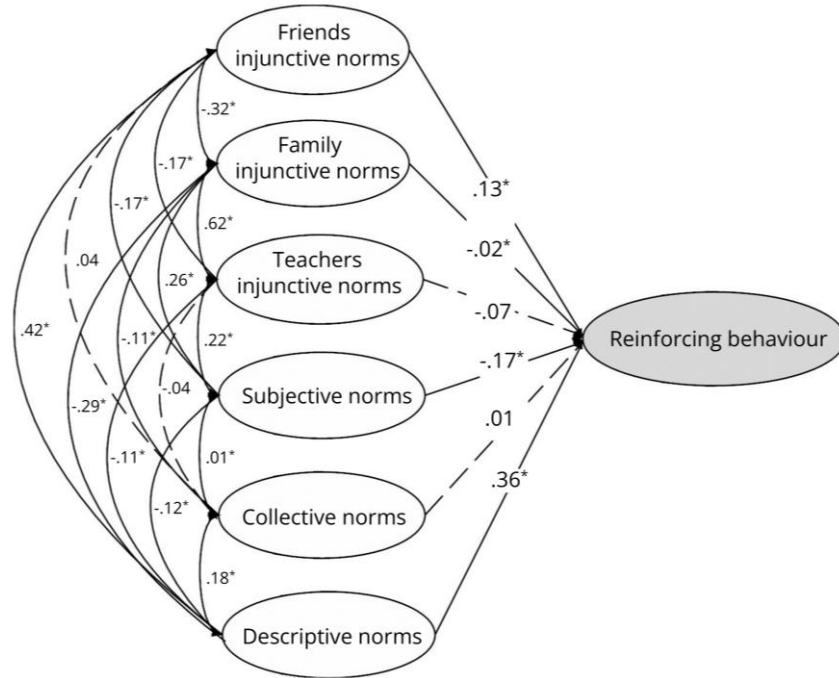
Note. * $p < .05$. Friends' injunctive norms reflect approval of online risky behaviour, while the family and teachers' injunctive norms reflect disapproval of such behaviours.

Figure 3. Graphical Representation of the Structural Equation Model (SEM) for Passive Behaviour.



Note. * $p < .05$. Friends' injunctive norms reflect approval of online risky behaviour, while the family and teachers' injunctive norms reflect disapproval of such behaviours.

Figure 4. Graphical Representation of the Structural Equation Model (SEM) for Reinforcing Behaviour.



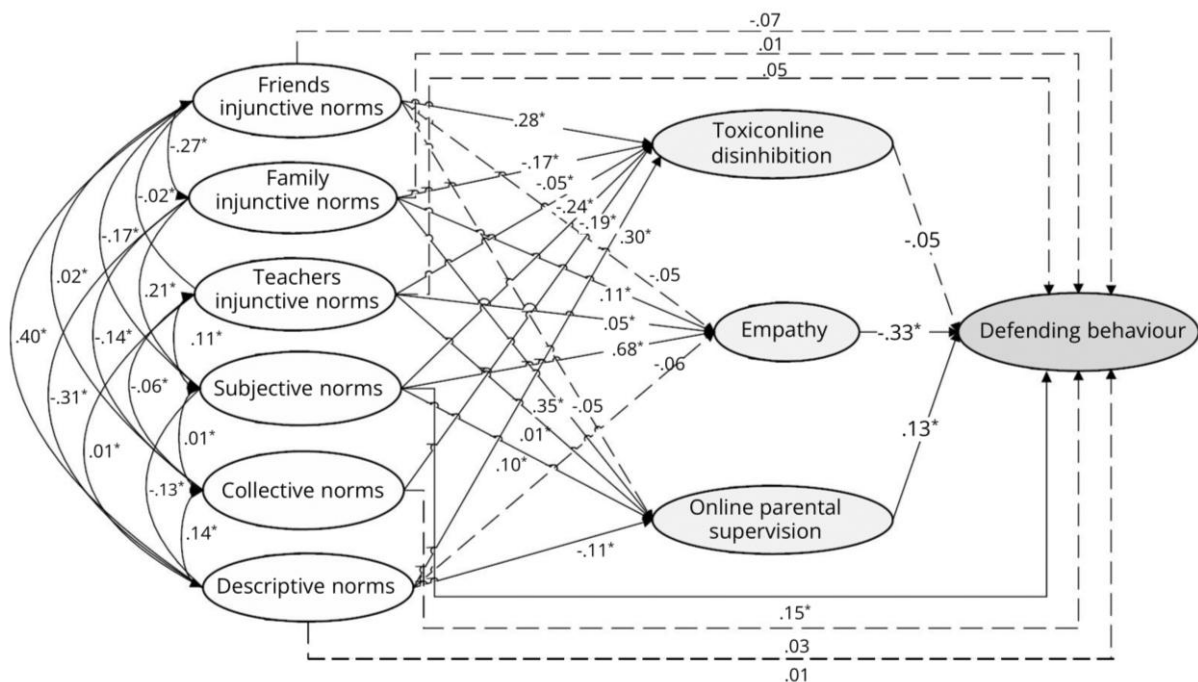
Note. $*p < .05$. Friends' injunctive norms reflect approval of online risky behaviour, while the family and teachers' injunctive norms reflect disapproval of such behaviours.

TNSB Model With the Inclusion of Toxic Online Disinhibition, Empathy and Online Parental Supervision

Subsequently, and following the third objective, toxic online disinhibition, empathy, and online parental supervision were included as additional variables in the model.

The model for the defending behaviour obtained a Mardia's Coefficient (Mardia, 1970) of 503.1387 and the fit indices: χ^2 S-B (821) = 8604.5437, $p < .001$, RMSEA = .04, 90% CI [.037, .040], SRMR = .06, CFI = .98, NNFI = .98, with an explained variance of 31.8%. The defending behaviour showed a significant and positive relation with subjective norms ($\beta = .15$), empathy ($\beta = .33$) and online parental supervision ($\beta = .13$; see Figure 5).

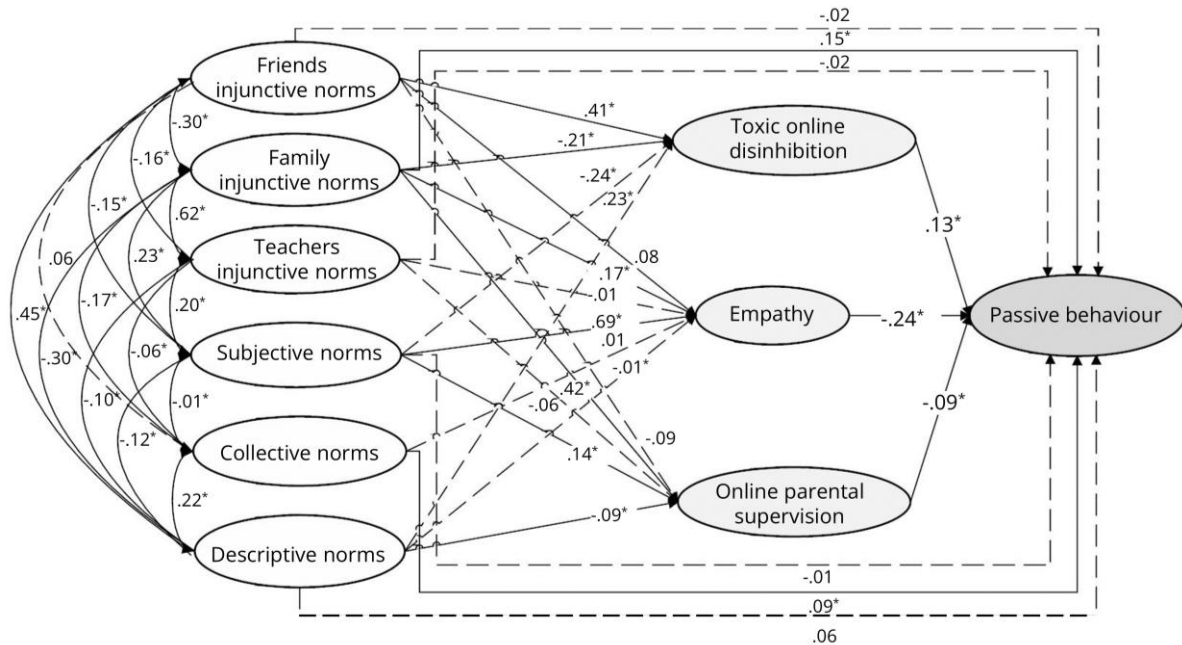
Figure 5. Graphical Representation of the Structural Equation Model (SEM) With Additional Variables for Defending Behaviour.



Note. $*p < .05$. Friends' injunctive norms reflect approval of online risky behaviour, while the family and teachers' injunctive norms reflect disapproval of such behaviours.

The model for the passive behaviour obtained a Mardia's Coefficient (Mardia, 1970) of 506.1718 and the fit indices: χ^2 S-B (780) = 3445.7049, $p < .001$, RMSEA = .04, SRMR = .04, 90% CI [.036, .039], CFI = .98, NNFI = .97, with an explained variance of 12.6%. Passive behaviour showed significant and positive relations with family injunctive norms ($\beta = .15$), collective injunctive norms ($\beta = .09$) and toxic online disinhibition ($\beta = .13$; see Figure 6). In contrast, it was negatively associated with empathy ($\beta = -.24$) and online parental supervision ($\beta = -.09$).

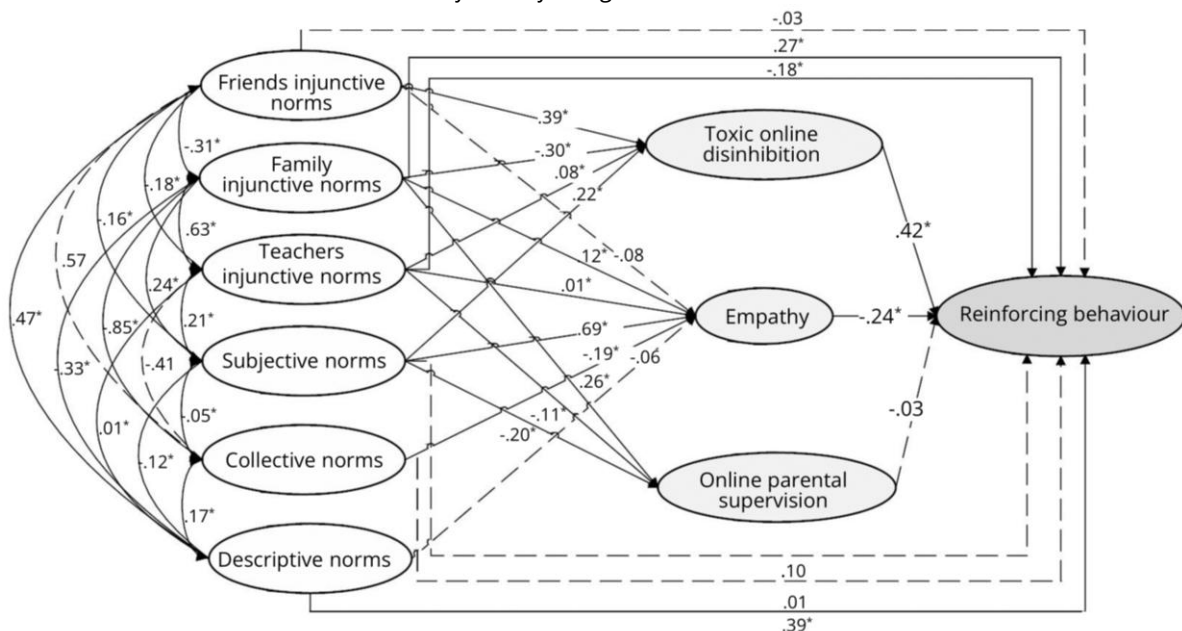
Figure 6. Graphical Representation of the Structural Equation Model (SEM) With Additional Variables for Passive Behaviour.



Note. * $p < .05$. Friends' injunctive norms reflect approval of online risky behaviour, while the family and teachers' injunctive norms reflect disapproval of such behaviours.

The model for the reinforcing behaviour obtained a Mardia's Coefficient (Mardia, 1970) of 638.1712 and the fit indices: χ^2 S-B (782) = 3841.0297, $p < .001$, RMSEA = .04, 90% CI [.041, .044], SRMR = .04, CFI = .97, NNFI = .97, with an explained variance of 40.1%. The reinforcing behaviour showed significant positive associations with descriptive norms ($\beta = .39$), family injunctive norms ($\beta = .27$), and toxic online disinhibition ($\beta = .42$). Additionally, negative associations were obtained with teacher injunctive norms ($\beta = -.18$) and empathy ($\beta = -.24$; see Figure 7). The coefficients of the three models are presented in Table 3.

Figure 7. Graphical Representation of the Structural Equation Model (SEM) With Additional Variables for Reinforcing Behaviour.



Note. $^*p < .05$. Friends' injunctive norms reflect approval of online risky behaviour, while the family and teachers' injunctive norms reflect disapproval of such behaviours.

Table 3. *Coefficients and R-Squared Results of the Structural Equation Modeling of the Three Bystander Behaviours.*

	Defending behaviour		Passive behaviour		Reinforcing behaviour	
	Model 1 β (SE)	Model 2 β (SE)	Model 1 β (SE)	Model 2 β (SE)	Model 1 β (SE)	Model 2 β (SE)
Friends' injunctive norms	-.11* (.05)	-.07 (.05)	.06 (.05)	-.02 (.06)	.13* (.03)	-.03 (.03)
Family injunctive norms	.05 (.05)	.01 (.06)	.10* (.05)	.15* (.06)	-.02* (.03)	.27* (.05)
Teachers' injunctive norms	.07 (.05)	.05 (.03)	-.08 (.04)	-.02 (.03)	-.07 (.02)	-.18* (.02)
Subjective norms	.39* (.02)	.15* (.03)	-.22* (.02)	-.01 (.03)	-.17* (.01)	.10 (.01)
Collective norms	.01 (.12)	.03 (.14)	.10* (.12)	.09* (.12)	.01 (.03)	.01 (.05)
Descriptive norms	-.05 (.05)	.01 (.06)	.12* (.05)	.06 (.05)	.36* (.04)	.39* (.04)
Toxic online disinhibition	—	-.05 (.07)	—	.13* (.03)	—	.42* (.04)
Empathy	—	.33* (.03)	—	-.24* (.07)	—	-.24* (.01)
Online parental supervision	—	.13* (.03)	—	-.09* (.02)	—	-.03 (.01)
R^2	.233	.318	.083	.126	.269	.401

Note. $^*p < .05$. Model 1 represents the TNSB (social norms), and Model 2 represents the TNSB with the additional variables (toxic online disinhibition, empathy and online parental supervision).

Discussion

The present study aimed to understand adolescent bystanders' behaviours to cyberhate, to test whether the TNSB can be useful as an explanatory framework of their behaviours, and to examine whether the TNSB shows a better explanatory capacity with the inclusion of behavioural attribute (i.e., toxic online disinhibition), personal (i.e., empathy) and contextual (i.e., online parental supervision) variables.

The findings obtained concerning the first objective show that adolescents mainly opt for a defending behaviour when they see an episode of cyberhate, which supports hypothesis H1. These results align with previous research, which pointed out that youths and adolescents tend to intervene when they are bystanders in this phenomenon (Costello et al., 2023; Obermaier, 2024). Furthermore, it is noteworthy that the highest scores of adolescents are those regarding behaviours that involve seeking social support, such as telling a trusted adult or encouraging the victim to do so. This may reflect a tendency toward help-seeking strategies, which is consistent with previous studies on bystander behaviour to hate speech, which have found that individuals with a defender profile are characterised by high levels of help-seeking behaviour (Wachs, Wettstein, et al., 2024). These results suggest that adolescents perceive social support-seeking as the most effective strategy to counteract cyberhate, which is consistent with previous research on cyberbullying that highlights it as such (Mora-Merchán et al., 2021). However, it is crucial to highlight that reinforcing behaviours, and especially passive ones, are relatively common, so it is essential to understand what leads adolescents to adopt these behaviours rather than defending ones.

In this sense, the results of the second research objective show that the TNSB is a good explanatory theory for adolescents' behaviour and, in this case, to explain their behaviour as bystanders of cyberhate, as proposed in hypothesis 2 (H2a and H2b). Specifically, active behaviours (i.e., defending and reinforcing) obtained greater explanatory capacity, which is understandable considering that the TNSB is postulated to explain behaviour, not its absence (Rimal & Real, 2005; Rimal & Yilma, 2022). Nevertheless, the results have also shown explanatory power for passive bystander behaviour, reinforcing this theory's potential.

Among the social norms that have been relevant in explaining bystanders' behaviour towards the phenomenon, the influence of subjective norms has been highlighted. Subjective norms were directly related to all three types of behaviours, being especially influential in defending. This suggests that adolescents value what people who

are relevant to them expect from their online behaviour (Batool & Lewis, 2022). Thus, adolescents, anticipating that people significant to them would expect them to defend the victim when facing an episode of cyberhate, feel more supported and motivated to intervene in such situations. Likewise, they may feel a sense of responsibility not to deviate from these social expectations (Lledó Rando et al., 2023) and are, therefore, less likely to respond in a passive or reinforcing way.

Under this premise, friends' injunctive norms were also relevant in explaining defending and reinforcing behaviour. In this sense, adolescents who perceive that their friends accept and normalise violent behaviour on the Internet are more likely to act as reinforcers, and less likely to act as defenders. It suggests that perceived social approval within the peer group may serve as a key mechanism through which individuals evaluate and justify their own behaviour to cyberhate. This finding expands on previous literature by showing that peer norms about online hateful behaviour are not only associated with cyberhate perpetration (Wachs, Wettstein, et al., 2022) but also influence how adolescents respond as bystanders when witnessing such incidents (Wachs, Castellanos, et al., 2023). In particular, when peer groups convey a permissive or tolerant attitude toward cyberhate, this may lower adolescents' perceived moral obligation to intervene or reject such behaviours, thereby reducing defending behaviours and fostering reinforcing behaviours. These results highlight the central role of peer influence in shaping bystander behaviour, consistent with evidence from other areas of cyberviolence, such as cyberbullying, where peers and the social context have been identified as key predictors of intervention (Domínguez-Hernández et al., 2018).

Similarly, for the reinforcing and passive behaviour, another relevant figure in their social environment seems to be the family. Adolescents who felt their family accepted or normalised cyberviolence were more likely to reinforce it, while those who thought their family would disapprove were more likely to stay passive. The family is responsible for teaching their children values, social norms and accepted behaviours (Livingstone & Byrne, 2018). Therefore, adolescents' attitudes and behaviours may reflect the internalisation of these family teachings, either by endorsing or adopting a passive attitude toward cyberviolence. This last case may be because, despite not accepting cyberviolence, families tend to encourage the avoidance of risky online behaviours (Wright et al., 2021). In this sense, families could adopt a no-action perspective when adolescents are bystanders, teaching them to avoid conflict and encouraging passive behaviour to reduce potential risks.

Furthermore, descriptive norms played a significant role in both passive and reinforcing behaviours. Additionally, in the context of passive behaviour, collective norms were also relevant. These norms refer, respectively, to perceptions of the prevalence of cyberhate and the presence of cyberhate in adolescents' social environment (Chung & Rimal, 2016). Frequent contact with episodes of cyberhate may desensitise adolescents, leading them to perceive these situations as less severe or to believe that they are impossible to resolve due to their high frequency, which in turn decreases the likelihood that they will intervene when they see these behaviours (Rudnicki et al., 2023; Schmid et al., 2022). Additionally, a high perceived prevalence of cyberhate may lead adolescents to normalise these behaviours, as previous studies have shown that adolescents' perceptions of the frequency of how common hateful behaviour online is among their peers are positively linked to higher exposure to cyberhate (Turner et al., 2023). This normalization may lead to passive bystander behaviour or even adopting aggressive or supportive behaviours towards the aggressor (Wachs et al., 2021). This interpretation gains further relevance when considering the actual prevalence rates observed in our sample, where around the 80% of adolescents have seen cyberhate, more than a third reported having received hate messages in the past year, and nearly a quarter acknowledged having sent or searched for them. This not only underscores the ubiquity of the phenomenon but also suggests that a significant portion of bystanders may have prior personal experience with cyberhate—an experience that could potentially influence the normalisation of the behaviour and reduce the likelihood of intervention (Obermaier, 2024; Rudnicki et al., 2023).

Likewise, one possible explanation for the specific association between collective norms and passive behaviour is that perceiving cyberhate as widespread in one's environment might contribute to an unwillingness to intervene, potentially due to concerns about social repercussions or retaliation (Domínguez-Hernández et al., 2018). Moreover, the influence of collective norms tends to be more diffuse and indirect compared to the more immediate and emotionally salient norms of close referents such as friends or family (Geber et al., 2021). Consequently, while these broader norms may lack the strength to motivate active responses such as defending or reinforcing, they may contribute to passivity by normalising inaction or discouraging involvement. All these

results, which show the influence of subjective, friends and family norms, as well as descriptive and collective norms, underline the relevance of social norms in understanding behaviour, as they determine whether we consider some situations to be severe or socially unacceptable enough to intervene (Rudnicki et al., 2023).

Findings regarding the third objective of the present research indicate that including additional variables improved the explanatory capacity of TNSB, as hypothesised in hypothesis 3 (H3). It supports that behaviour is not solely a result of the normalisation of environmental behaviours but is also influenced by personal, contextual, and even behavioural factors (Rimal & Yilma, 2022). In this sense, empathy, as a personal factor, was crucial for all three types of bystander behaviours, especially for the defending one (H3a). Higher levels of empathy in adolescents promoted the adoption of defending behaviour and decreased passive and reinforcing behaviours. These findings align with previous studies demonstrating that empathy fosters a willingness to intervene as a bystander to cyberhate (Costello et al., 2023; Wachs, Krause, et al., 2023). However, results have also been obtained regarding its negative relation with reinforcing behaviours, which is a significant contribution. This could be due to empathy being inversely related to the perpetration and acceptance of cyberhate (Celuch et al., 2022; Wachs, Bilz, et al., 2022), implying that it may also influence reinforcing aggressive behaviours. These results highlight the importance of including the development of socio-emotional competence within the online context in future psychoeducational programs.

The relevance of online socio-emotional development becomes more evident considering that toxic online disinhibition has also been found to be relevant for being passive and especially for being reinforcing (H3b). Adolescents who feel more disinhibited in online environments have a greater tendency to be passive and to reinforce hateful behaviours. Decreased social constraints and inhibitions, common in face-to-face interactions, cause adolescents to perceive themselves as part of an anonymous crowd, which decreases their sense of responsibility (Suler, 2004). Perceived responsibility increases the likelihood of intervening in cases of cyberhate (Obermaier, 2024), so online disinhibition could contribute to greater passivity in bystander behaviour by reducing this sense of responsibility. Also, this disinhibition caused by anonymity, the feeling of lack of rules and consequences, asynchrony, or lack of eye contact makes people not see cyberhate episodes as severe and favour aggressive behaviours (Wachs & Wright, 2018), and in the case of bystanders, reinforcing behaviours.

Finally, online parental supervision was related to being a defender and decreasing passive bystander behaviours (H3c). This may be explained by parental mediation strategies increasing children's awareness of online risks, fostering the internalisation of safety guidelines, and strengthening their ability to cope with online dangers such as cyberhate (Wright et al., 2021). Furthermore, although adolescents place great importance on peer influence, the family remains a key pillar as a protector and provider of guidance in online environments (Livingstone & Byrne, 2018). Therefore, this online parental supervision could also help mitigate other factors that inhibit intervention, such as friends' social norms. Thus, these findings underline the importance of involving families in cyberhate prevention programs, recognising this phenomenon as a challenge that requires the collective effort of the whole community.

Practical Implications

The findings of this study offer relevant insights for the development of interventions aimed at preventing cyberhate and fostering defending bystander behaviours among adolescents. Since defending behaviours, particularly help-seeking, were the most appreciated by participants, prevention programs should prioritise encouraging adolescents to seek support from trusted people when witnessing cyberhate. This approach is especially important considering the influence of subjective norms, as adolescents are highly responsive to the expectations of their social environment. Reinforcing defending behaviours as socially valued and expected while fostering supportive cultures may enhance adolescents' willingness to intervene.

Such supportive cultures should be especially promoted among peer groups. Given the strong influence of peer norms on bystander behaviour, interventions should focus on creating group dynamics that reject cyberhate and encourage defending actions. Peer-led initiatives and mentoring programs can be particularly effective in reshaping group norms and reinforcing prosocial behaviour.

Family involvement also plays a crucial role in guiding adolescents' behaviours towards cyberhate. It is important to emphasise to families the need to express disapproval of online violence clearly and to communicate expectations that their children act as defenders. Promoting positive online behaviour at home through open dialogue and active supervision can help create a supportive context that reinforces defending behaviours.

In addition, the importance of empathy in predicting defending behaviours highlights the need to strengthen adolescents' socio-emotional competencies. Intervention programmes should incorporate educational strategies aimed at developing empathy and emotional intelligence, as these abilities foster defending behaviours and help reduce passive or reinforcing bystander behaviours.

Lastly, considering the results about toxic online disinhibition, it is essential to educate adolescents about the real consequences of engaging with or tolerating cyberhate. Addressing the psychological distance often created by digital anonymity requires promoting digital responsibility and ethical online behaviour. Educational programs should help adolescents understand that cyberhate causes real harm to individuals and communities, even if it happens behind a screen. For instance, classroom activities such as role-playing scenarios, guided discussions about the emotional impact of cyberhate, or analysis of real cases in the media can foster empathy and critical reflection.

Incorporating these practical implications into interventions and programs can effectively address cyberhate and encourage adolescents to adopt more proactive and supportive behaviours, ultimately creating a safer and more positive online environment.

Limitations and Future Lines of Research

Despite the contributions of this study, some limitations should be acknowledged, which also offer opportunities for future research. First, although the sample included schools with diverse socioeconomic backgrounds and public and semi-private institutions, it was selected through convenience sampling and is, therefore, not representative of the broader adolescent population. Moreover, the schools that agreed to participate may have been more sensitised to cyberviolence, potentially influencing students' behaviours. Future studies should aim to replicate these findings with representative and/or randomly selected samples to improve the generalizability of the results. Second, following the adaptation of the Theory of Normative Social Behaviour (TNSB) by Rimal and Yilma (2022), the study did not distinguish between actual behaviours and behavioural intentions, which may have influenced the interpretation of the results. Future research should differentiate between intentions and enacted behaviours to better capture the dynamics of bystander behaviours.

Additionally, the analysis did not differentiate between adolescents who were solely bystanders and those who had other roles in cyberhate, such as victims or perpetrators. Future studies could segment participants according to their level of involvement in cyberhate to examine whether social norms influence bystander behaviour differently across groups. Moreover, the cross-sectional nature of the study prevents drawing causal conclusions about the relationships between the variables. Longitudinal designs would allow for testing mediational pathways and evaluating the indirect effects of variables such as empathy, parental supervision, or social norms on defending or reinforcing behaviours. Future research would also be valuable in integrating all three types of bystander behaviours—defending, passive, and reinforcing—into a single model to analyse how these behaviours interact or coexist within individuals. Additionally, applying the TNSB framework to other forms of cyberviolence (e.g., cyberbullying, online harassment, or sexting) would help assess the theory's broader applicability and explanatory power in various online contexts.

Conclusion

The present study's findings contribute significantly to the understanding of adolescents' cyberhate behaviour and the empirical validation of the TNSB. The results revealed that adolescents tend to adopt defending behaviour to cyberhate, followed by passive and reinforcing behaviours. In understanding their decision to intervene in one way or another, the TNSB proved to be a useful theory to explain adolescents' cyberhate behaviour, especially when toxic online disinhibition, empathy and parental supervision were included as additional variables. These findings underline the importance of social norms, along with behavioural, personal

and contextual variables, in understanding cyberhate bystanders' behaviour, highlighting the need to address cyberhate as a community problem that requires a collective response.

Conflict of Interest

The authors have no conflicts of interest to declare.

Use of AI Services

The authors declare they have not used any AI services to generate or edit any part of the manuscript or data.

Authors' Contribution

Olga Jiménez-Díaz: conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft, visualization. **Joaquín A. Mora-Merchán:** resources, writing—review & editing, investigation, supervision, project administration, funding acquisition. **Paz Elipe:** conceptualization, writing—review & editing, supervision. **Rosario Del Rey:** resources, writing—review & editing, investigation, supervision, project administration, funding acquisition.

Acknowledgement

This publication is part of the R&D&I project PID2020-115913GB-I00, funded by MICIU/AEI/10.13039/501100011033/ and the Grant FPU21/05405 by the Ministry of Universities of the Government of Spain.

Data Availability Statement

The data will be provided upon request.

Ethics Approval Statement

This study was approved by the Andalusian Biomedical Research Ethics Coordinating Committee (2563-N-20).

References

- Álvarez-García, D., Thornberg, R., & Suárez-García, Z. (2021). Validation of a scale for assessing bystander responses in bullying. *Psicothema*, 33(4), 623–630. <https://doi.org/10.7334/psicothema2021.140>
- Bastiaensens, S., Pabian, S., Vandebosch, H., Poels, K., Van Cleemput, K., DeSmet, A., & De Bourdeaudhuij, I. (2016). From normative influence to social pressure: How relevant others affect whether bystanders join in cyberbullying. *Social Development*, 25(1), 193–211. <https://doi.org/10.1111/sode.12134>
- Batool, S. S., & Lewis, C. A. (2022). Does positive parenting predict pro-social behavior and friendship quality among adolescents? Emotional intelligence as a mediator. *Current Psychology*, 41(4), 1997–2011. <https://doi.org/10.1007/s12144-020-00719-y>
- Bauman, S., Perry, V. M., & Wachs, S. (2021). The rising threat of cyberhate for young people around the globe. In M. F. Wright & L. B. Schiamberg (Eds.), *Child and adolescent online risk exposure: An ecological perspective* (pp. 149–175). Elsevier. <https://doi.org/10.1016/B978-0-12-817499-9.00008-9>
- Bedrosova, M., Mylek, V., Dedkova, L., & Velicu, A. (2023). Who is searching for cyberhate? Adolescents' characteristics associated with intentional or unintentional exposure to cyberhate. *Cyberpsychology, Behavior, and Social Networking*, 26(7), 462–471. <https://doi.org/10.1089/cyber.2022.0201>
- Blaya, C., & Audrin, C. (2019). Toward an understanding of the characteristics of secondary school cyberhate perpetrators. *Frontiers in Education*, 4, Article 46. <https://doi.org/10.3389/feduc.2019.00046>

- Bührer, S., Koban, K., & Matthes, J. (2024). The WWW of digital hate perpetration: What, who, and why? A scoping review. *Computers in Human Behavior*, 159, Article 108321. <https://doi.org/10.1016/j.chb.2024.108321>
- Celuch, M., Oksanen, A., Räsänen, P., Costello, M., Blaya, C., Zych, I., Llorent, V. J., Reichelmann, A., & Hawdon, J. (2022). Factors associated with online hate acceptance: A cross-national six-country study among young adults. *International Journal of Environmental Research and Public Health*, 19(1), Article 534. <https://doi.org/10.3390/ijerph19010534>
- Chung, A., & Rimal, R. N. (2016). Social norms: A review. *Review of Communication Research*, 4, 1–28. <https://doi.org/10.12840/issn.2255-4165.2016.04.01.008>
- Costello, M., Hawdon, J., & Ratliff, T. N. (2017). Confronting online extremism: The effect of self-help, collective efficacy, and guardianship on being a target for hate speech. *Social Science Computer Review*, 35(5), 587–605. <https://doi.org/10.1177/0894439316666272>
- Costello, M., Hawdon, J., Reichelmann, A. V., Oksanen, A., Blaya, C., Llorent, V. J., Räsänen, P., & Zych, I. (2023). Defending others online: The influence of observing formal and informal social control on one's willingness to defend cyberhate victims. *International Journal of Environmental Research and Public Health*, 20(15), Article 6506. <https://doi.org/10.3390/ijerph20156506>
- Domínguez-Hernández, F., Bonell, L., & Martínez-González, A. (2018). A systematic literature review of factors that moderate bystanders' actions in cyberbullying. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 12(4), Article 1. <https://doi.org/10.5817/CP2018-4-1>
- Garland, J., Ghazi-Zahedi, K., Young, J.-G., Hébert-Dufresne, L., & Galesic, M. (2022). Impact and dynamics of hate and counter speech online. *EPJ Data Science*, 11, Article 3. <https://doi.org/10.1140/epjds/s13688-021-00314-6>
- Geber, S., Baumann, E., Czerwinski, F., & Klimmt, C. (2021). The effects of social norms among peer groups on risk behavior: A multilevel approach to differentiate perceived and collective norms. *Communication Research*, 48(3), 319–345. <https://doi.org/10.1177/0093650218824213>
- Hawdon, J., Reichelmann, A., Costello, M., Llorent, V. J., Räsänen, P., Zych, I., Oksanen, A., & Blaya, C. (2023). Measuring hate: Does a definition affect self-reported levels of perpetration and exposure to online hate in surveys? *Social Science Computer Review*, 42(3), 812–831. <https://doi.org/10.1177/08944393231211270>
- Hayashi, Y., & Tahmasbi, N. (2022). Psychological predictors of bystanders' intention to help cyberbullying victims among college students: An application of theory of planned behavior. *Journal of Interpersonal Violence*, 37(13–14), NP11333–NP11357. <https://doi.org/10.1177/0886260521992158>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jiménez-Díaz, O., & Del Rey, R. (2025). Cyberhate in adolescents and youths: A systematic review of labels and associated factors. *Aggression and Violent Behavior*, 81, Article 102023. <https://doi.org/10.1016/j.avb.2024.102023>
- Jiménez-Díaz, O., Del Rey, R., Espino, E., & Casas, J. A. (2025). Understanding cyberhate: the importance of like-seeking behaviours/Comprendiendo el ciberodio: la importancia de los comportamientos de búsqueda de likes. *Journal for the Study of Education and Development*, 48(3), 601–629. <https://doi.org/10.1177/02103702251356207>
- Jiménez-Díaz, O., Wachs, S., Del Rey, R., & Mora-Merchán, J. A. (2024). Associations between searching and sending cyberhate: The moderating role of the need of online popularity and toxic online disinhibition. *Cyberpsychology, Behavior, and Social Networking*, 28(1), 37–43. <https://doi.org/10.1089/cyber.2024.0305>
- Livingstone, S., & Byrne, J. (2018). Parenting in the digital age: The challenges of parental responsibility in comparative perspective. In G. Mascheroni, C. Ponte, & A. Jorge (Eds.), *Digital parenting: The challenges for families in the digital age* (pp. 19–30). Nordicom. <https://urn.kb.se/resolve?urn=urn:nbn:se:norden:org:diva-12015>
- Livingstone, S., Mascheroni, G., Dreier, M., Chaudron, S., & Lagae, K. (2015). *How parents of young children manage digital devices at home: The role of income, education and parental style*. EU Kids Online, LSE. <https://eprints.lse.ac.uk/63378/>
- Lledó Rando, C., Perles Novas, F., & San Martín García, J. (2023). Exploring social norms and control coercive on intimate partner violence in the young. Implications for prevention. *Journal of Interpersonal Violence*, 38(9–10), 6695–6722. <https://doi.org/10.1177/08862605221137721>

- Llorent, V.-J., Seade-Mejía, C., & Vélez-Calvo, X. (2023). Lockdown, cyberhate, and protective factor of social-emotional and moral competencies in primary education. *Comunicar: Revista Científica de Comunicación y Educación*, 31(77), 109–118. <https://doi.org/10.3916/C77-2023-09>
- March, E., & Marrington, J. (2019). A qualitative analysis of internet trolling. *Cyberpsychology, Behavior and Social Networking*, 22(3), 192–197. <https://doi.org/10.1089/cyber.2018.0210>
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530. <https://doi.org/10.1093/biomet/57.3.519>
- Martin-Criado, J.-M., Casas, J.-A., Ortega-Ruiz, R., & Del Rey, R. (2021). Parental supervision and victims of cyberbullying: Influence of the use of social networks and online extimacy. *Revista de Psicodidáctica (English Ed.)*, 26(2), 160–167. <https://doi.org/10.1016/j.psicoe.2021.04.002>
- Menesini, E., & Salmivalli, C. (2017). Bullying in schools: The state of knowledge and effective interventions. *Psychology, Health & Medicine*, 22(sup1), 240–253. <https://doi.org/10.1080/13548506.2017.1279740>
- Mora-Merchán, J. A., Espino, E., & Del Rey, R. (2021). Development of effective coping strategies to reduce school bullying and their impact on stable victims. *Psychology, Society & Education*, 13(3), 55–66. <https://doi.org/10.25115/psyse.v13i3.5586>
- Obermaier, M. (2024). 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., Schmuck, D., & Saleem, M. (2023). I'll be there for you? Effects of Islamophobic online hate speech and counter speech on Muslim in-group bystanders' intention to intervene. *New Media & Society*, 25(9), 2339–2358. <https://doi.org/10.1177/14614448211017527>
- Rimal, R. N., & Real, K. (2005). How behaviors are influenced by perceived norms: A test of the theory of normative social behavior. *Communication Research*, 32(3), 389–414. <https://doi.org/10.1177/0093650205275385>
- Rimal, R. N., & Yilma, H. (2022). Descriptive, injunctive, and collective norms: An expansion of the theory of normative social behavior (TNSB). *Health Communication*, 37(13), 1573–1580. <https://doi.org/10.1080/10410236.2021.1902108>
- Rudnicki, K., Vandebosch, H., Voué, P., & Poels, K. (2023). Systematic review of determinants and consequences of bystander interventions in online hate and cyberbullying among adults. *Behaviour & Information Technology*, 42(5), 527–544. <https://doi.org/10.1080/0144929X.2022.2027013>
- Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior*, 22(1), 1–15. [https://doi.org/10.1002/\(SICI\)1098-2337\(1996\)22:1<1::AID-AB1>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-2337(1996)22:1<1::AID-AB1>3.0.CO;2-T)
- Sasson, H., & Mesch, G. (2016). Gender differences in the factors explaining risky behavior online. *Journal of Youth and Adolescence*, 45(5), 973–985. <https://doi.org/10.1007/s10964-016-0465-7>
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514. <https://doi.org/10.1007/BF02296192>
- Schmid, U. K., Kümpel, A. S., & Rieger, D. (2022). How social media users perceive different forms of online hate speech: A qualitative multi-method study. *New Media & Society*, 26(5), 2614–2632. <https://doi.org/10.1177/14614448221091185>
- Staub, E. (2005). The origins and evolution of hate, with notes on prevention. In R. J. Sternberg (Ed.), *The psychology of hate* (pp. 51–66). American Psychological Association. <https://doi.org/10.1037/10930-003>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>
- Thornberg, R., & Jungert, T. (2013). Bystander behavior in bullying situations: Basic moral sensitivity, moral disengagement and defender self-efficacy. *Journal of Adolescence*, 36(3), 475–483. <https://doi.org/10.1016/j.adolescence.2013.02.003>
- Turner, N., Holt, T. J., Brewer, R., Cale, J., & Goldsmith, A. (2023). Exploring the relationship between opportunity and self-control in youth exposure to and sharing of online hate content. *Terrorism and Political Violence*, 35(7), 1604–1619. <https://doi.org/10.1080/09546553.2022.2066526>

- Udris, R. (2014). Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale. *Computers in Human Behavior*, 41, 253–261. <https://doi.org/10.1016/j.chb.2014.09.036>
- Van Ouytsel, J., Ponnet, K., Walrave, M., & d'Haenens, L. (2017). Adolescent sexting from a social learning perspective. *Telematics and Informatics*, 34(1), 287–298. <https://doi.org/10.1016/j.tele.2016.05.009>
- Voggeser, B. J., Singh, R. K., & Göritz, A. S. (2018). Self-control in online discussions: Disinhibited online behavior as a failure to recognize social cues. *Frontiers in Psychology*, 8, Article 2372. <https://doi.org/10.3389/fpsyg.2017.02372>
- Wachs, S., Bilz, L., Wettstein, A., & Espelage, D. L. (2024). Validation of the multidimensional bystander responses to racist hate speech scale and its association with empathy and moral disengagement among adolescents. *Aggressive Behavior*, 50(1), Article e22105. <https://doi.org/10.1002/ab.22105>
- 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., Castellanos, M., Wettstein, A., Bilz, L., & Gámez-Guadix, M. (2023). Associations between classroom climate, empathy, self-efficacy, and countering hate speech among adolescents: A multilevel mediation analysis. *Journal of Interpersonal Violence*, 38(5–6), 5067–5091. <https://doi.org/10.1177/08862605221120905>
- 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(6), 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, 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., Wettstein, A., Bilz, L., Krause, N., Ballaschk, C., Kansok-Dusche, J., & Wright, M. F. (2022). Playing by the rules? An investigation of the relationship between social norms and adolescents’ hate speech perpetration in schools. *Journal of Interpersonal Violence*, 37(21–22), NP21143–NP21164. <https://doi.org/10.1177/08862605211056032>
- Wachs, S., & Wright, M. F. (2018). Associations between bystanders and perpetrators of online hate: The moderating role of toxic online disinhibition. *International Journal of Environmental Research and Public Health*, 15(9), Article 2030. <https://doi.org/10.3390/ijerph15092030>
- Wright, M. F., Wachs, S., & Gámez-Guadix, M. (2021). Youths’ coping with cyberhate: Roles of parental mediation and family support. *Comunicar*, 29(67), 21–33. <https://doi.org/10.3916/C67-2021-02>

About Authors

Olga Jiménez-Díaz is a PhD candidate in Psychology who has a predoctoral contract below the University Teacher Training Program by the Ministry of Universities of the Government of Spain [Code: FPU21/05405]. She is a member of the IASED research group at Universidad de Sevilla (Spain). Her line of research focuses on understanding the phenomenon of cyberhate in adolescents and youths.

<https://orcid.org/0000-0003-1896-2737>

Joaquín A. Mora-Merchán is Associate Professor in the Department of Developmental and Educational Psychology and a member of the IASED research group at Universidad de Sevilla (Spain). His career has been focused, above all, on the study of school violence problems, especially bullying and cyberbullying, and the ways to deal with them both individually and from educational institutions. His work has been published in national and international journals, and she has presented her contributions at prestigious congresses and conferences.

<https://orcid.org/0000-0002-8558-6965>

Paz Elipe is an Associate Professor in the Department of Psychology at the University of Jaén. Her research trajectory has been focused on the emotional impact of bullying and cyberbullying phenomena and, recently, on the study of LGBTQ+ bullying. She is a member of the IASED research group, and she actively participates in different regional, state, and international research projects.

<https://orcid.org/0000-0001-5210-3747>

Rosario Del Rey is a Professor of Developmental and Educational Psychology and principal investigator of the IASED research group at Universidad de Sevilla (Spain). She is currently co-president of the International Observatory for School Climate and Violence Prevention (IOSCVP). Her lines of research focus on school coexistence, bullying, cyberbullying, and sexting.

<https://orcid.org/0000-0002-1907-5489>

Correspondence to

Rosario Del Rey, Department of Developmental and Educational Psychology, Universidad de Sevilla, Camilo José Cela Street, 41013, Seville, Spain, delrey@us.es

© Author(s). The articles in Cyberpsychology: Journal of Psychosocial Research on Cyberspace are open access articles licensed under the terms of the [Creative Commons BY-SA 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/) which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited and that any derivatives are shared under the same license.