






# At the core of cyberaggression: A group-based explanation

Elena Trifiletti<sup>1</sup>  | Marco Giannini<sup>1</sup> | Loris Vezzali<sup>2</sup>  | Soraya E. Shamloo<sup>1,2</sup>  |  
Martina Faccini<sup>1</sup>  | Veronica M. Cocco<sup>3</sup> 

<sup>1</sup>Department of Human Sciences, University of Verona, Verona, Italy

<sup>2</sup>Faculty of Medicine, University of Modena and Reggio Emilia, Modena, Italy

<sup>3</sup>Department of Humanities, Social Sciences and Cultural Industries, University of Parma, Parma, Italy

## Correspondence

Soraya E. Shamloo, Faculty of Medicine, University of Modena and Reggio Emilia, Largo del Pozzo, 71, 41125 Modena, Italy.  
Email: [soraya.shamloo@gmail.com](mailto:soraya.shamloo@gmail.com)

## Abstract

In this study, prototypicality of the aggressor was tested as a group-level factor predicting social media users' active participation in cyberaggression. Participants were exposed to a fictitious conversation, in which either a prototypical versus non-prototypical user posted an aggressive comment as a reply to a provocative comment. In line with self-categorization theory, we hypothesized that bystander participants would post an aggressive comment and rate the aggression as acceptable to a greater extent in the prototypical than in the non-prototypical condition. Furthermore, we predicted that perceived normativity of aggression would mediate the effect of prototypicality. Results supported these predictions and showed that prototypical members affect the extent to which collective aggressive behaviors in online interactions are approved and enacted. These findings highlight the importance of group-level factors in the study of cyberaggression and provide important information for understanding the psychological underpinnings of collective forms of online aggression.

## KEYWORDS

bystanders, cyberaggression, prototypicality, self-categorization theory, social media

## 1 | INTRODUCTION

In recent years, scholars of various disciplines have devoted their attention to the spread of attitudes and behaviors across social networks (Christakis & Fowler 2013; Fagan et al., 2007; Powell et al., 2015). Research has shown that social contagion occurs for some physical and psychological conditions, such as obesity (Christakis & Fowler, 2007; Powell et al., 2015), influenza (Christakis & Fowler, 2010), loneliness, and depression (Cacioppo et al., 2009; Rosenquist et al., 2011), as well as for a range of behaviors, including smoking (Christakis & Fowler, 2008), cooperation (Jordan et al., 2013), and physical exercise (Aral & Nicolaides, 2017). Extant literature suggests that social contagion effects are also observed for aggressive behavior. The aim of the present study was to examine this premise in the context of online aggression.

### 1.1 | Personal and situational factors underlying the spread of online aggression

Concerning the spread of aggressive behaviors, a recent study has revealed that US adolescents who had friends engaging in violent behaviors were more likely to do the same; notably, this association was not confined to immediate friends but extended to friends of friends (Bond & Bushman, 2017). Similarly, Green et al. (2017) demonstrated that gunshot violence is transmitted through networks of people. In information technology interactions, the spread of aggression can be more rapid and involve a larger number of people compared to face-to-face interactions. This can be easily observed in online dynamics such as online firestorms (Pfeffer et al., 2014), swearing (Kwon & Gruzd, 2017), or flaming on YouTube (Moor et al., 2010). Further, Yokotani and Takano (2021) showed that higher

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Aggressive Behavior* published by Wiley Periodicals LLC

rates of cyberbullying perpetrators in one's online social network were associated with greater likelihood of becoming a cyberbullying perpetrator. The dynamics of social interactions and communication in social media may be similar to those occurring in offline contexts, but they are amplified by technical artifacts of information technology (Pfeffer et al., 2014), such as speed, unrestrained flow of communication, and higher level of homogeneity within a multitude of interconnected network clusters. Apart from these structural factors, research has mainly focused on individual-level factors to explain participation in online aggression (see, e.g., Bastiaensens et al., 2014; Rost et al., 2016). Moving from the assumption that collective participation in online aggression is at least in part a group behavior (Tajfel, 1978), in the present work we aimed at investigating the role of prototypicality of the aggressor as a group-level situational factor affecting participation in online aggression.

Scholars have devoted a great deal of attention to the phenomenon of online aggression. Most studies in this field were concerned with cyberbullying among adolescents. Identifying the personal profiles of cyberbullies and their victims (e.g., Aoyama et al., 2011; Calvete et al., 2010; Vandebosch & Van Cleemput, 2009) has been a major focus of this line of research. Personal characteristics such as gender, age, personality, motivation, beliefs, and attitudes are amongst the factors most frequently associated with cyberbullying behavior (for a meta-analysis, see Kowalski et al., 2014). Studies on other forms of online disruptive behaviors have similarly highlighted the role of personal factors. For example, in two online studies, Buckels et al. (2014) found a significant association between trolling and sadism, psychopathy, and Machiavellianism. Situational factors have been only sporadically considered in the study of cyberbullying behaviors. Some scholars have analyzed the role of social and relational factors such as school climate and parental monitoring (see Kowalski et al., 2014), others have examined the technical features of online communication that may account for online bullying behaviors, such as anonymity (Moore et al., 2012), paucity of emotional cues, and ease of disseminating communication (Runions & Bak, 2015).

Research on online aggression has devoted considerably less attention to the spread of aggressive behaviors, failing to capture the group component of online communities. As observed in the studies on perpetration of cyberbullying or other aggressive behaviors, research in this field has primarily focused on personal factors. For instance, bystanders' active participation in cyberbullying has been associated with personal characteristics such as gender (Bastiaensens et al., 2014), empathy (Van Cleemput et al., 2014; Zych et al., 2019), moral disengagement (Wachs, 2012), and previous involvement in bullying (Barlińska et al., 2013; Wachs, 2012). Personal factors (e.g., intrinsic motivation, gender) have also been considered as determinants of collective participation in other forms of online aggression, such as online firestorms (Alonzo & Aiken, 2004; Rost et al., 2016). A few studies have analyzed situational determinants of bystanders' joining of cyberbullying, such as the context of violence (public vs. private; Barlińska et al., 2013) or friendship with other bystanders (Bastiaensens et al., 2014), but in general research on situational factors in this context has been scarce. Computer scientists have

pointed out that technical artifacts of social media communication such as speed, unrestrained flow of communication, and homogeneity of network clusters form the basis of massive participation in online aggression (see Pfeffer et al., 2014). These features may help to explain why aggressive (as well as nonaggressive) communication may spread so rapidly and reach a huge number of people but say little about when a single or a few aggressive messages will actually spread into the network. A situational factor that may shed light on this process is compliance with social norms. Bastiaensens et al. (2016) found that bystanders' decision to join in with cyberbullying is influenced by perceived approval by one's own friends. Similarly, Rösner and Krämer (2016) showed that more aggressive expressions were posted on a blog when peers' comments included aggressive wording.

## 1.2 | The spread of online aggression as a group process

In the present research, we argue that collective participation in online aggression should be regarded, at least in part, as a group process rather than solely a personal behavior. As such, research should investigate the role of situational group-level factors besides that of personal characteristics. According to Tajfel (1978), when uniform behavior is displayed by two or more individuals, that conduct may be thought of as a group behavior that is guided by social identity or categorization of the self as a group member (Turner, 1985). The adoption of individual-level factors to explain group processes and behaviors has been largely criticized (see Turner et al., 1987), because uniformity in behavior can be hardly explained in terms of individual differences, which should instead be better predictors of variability in individual behavior. Based on these considerations, we claim that group-level factors deserve more attention in the study of collective forms of online aggression, in which people tend to act in a uniform manner. In the present study, we aimed at investigating the role of prototypicality of the aggressor. We tested whether the decision to actively support an aggressor (i.e., to reproduce his/her behavior) by posting a hostile message depends on the aggressor's prototypicality (i.e., the extent to which the aggressor is prototypical of a salient social category). Furthermore, we investigated the perceived normativity of aggressive behavior as a potential mediator of the effect.

According to self-categorization theory (SCT; Turner et al., 1987), group processes can be explained in terms of depersonalization, namely the cognitive representation of the self and of other individuals in group terms. Factors that enhance the salience of an ingroup-outgroup categorization will increase perceptions of the self and of other individuals in terms of the stereotypical dimensions of group membership (Turner & Oakes, 1986). As representations of the self and others shift toward the group level, individual attitudes, beliefs, and behaviors will shift accordingly, conforming to the attitudes, beliefs, and behaviors displayed by fellow ingroup members. An important concept in SCT is that of prototypicality of group

members, which represents the extent to which group members are perceived to be similar to the group prototype. Group prototypes are “context specific [...] sets of attributes that define and prescribe attitudes, feelings, and behaviors that characterize one group and distinguish it from other groups” (Hogg, 2001, p. 187).

Initial research reported above (Bastiaensens et al., 2016; Rösner & Krämer, 2016) has highlighted the role of group norms as determinants of joint participation in online aggression. However, according to SCT group norms strictly depend upon prototypicality, and specifically on the position expressed by prototypical group members (Turner et al., 1987). For this reason, we focused on prototypicality of the aggressor and tested if an aggressor who is prototypical of a salient social category can contribute to shape a norm of aggressiveness and generate conformity to this norm. Following SCT, the extent to which individuals conform to group standards is a function of the relative persuasiveness of group members, that in turn is based on the extent to which their response is perceived to be prototypical of the ingroup. Highly prototypical ingroup members are able to exert considerable social influence on other group members, affecting the formation and adherence to group norms (Abrams et al., 1990; Hogg, 2001; Turner & Oakes, 1986). Therefore, people should be more inclined to conform to prototypical than to non-prototypical group members. In this study, this prediction was tested for the first time in an online setting, by exposing participants to the aggressive behavior of a prototypical versus non-prototypical social media user. We expected that aggression would be reproduced to a greater extent when aggressors are prototypical compared to when they are non-prototypical. A similar effect was expected for participants' perceptions of normativity regarding the aggressive behavior, which should be enhanced in response to a prototypical than to a non-prototypical aggressor and mediate the effect of prototypicality.

### 1.3 | The present research

The above predictions were tested in a laboratory experiment in which we manipulated prototypicality of the aggressor. Prototypicality was operationalized as the degree to which the online aggressor was representative of a belief, namely a positive opinion about integration of immigrants. It is worth noting that university students usually hold positive attitudes about immigrants (see, e.g., Boccato et al., 2015; Gravani et al., 2018; Vezzali & Giovannini, 2010; Vezzali et al., 2015), thus positive opinions about integration of immigrants are likely to activate an ingroup category.

Participants were presented with a blog entry reporting an article excerpt on the possible benefits of immigrants' integration and a comment section including comments of four fictitious users. Three of these users exchanged several comments in which they expressed a positive opinion about integration of immigrants. At the bottom of the comment section, we included a provocative comment allegedly posted by the fourth fictitious user (who opposed immigrants' integration) and an aggressive reply to this comment by one of the three initial users who was either prototypical or non-prototypical.

In the prototypical condition, the user was highly representative of the opinion expressed in the article excerpt (and shared by our participants, see below), while in the non-prototypical condition the user was less representative. Based on SCT assumptions, we predicted that participants would be more likely to post an aggressive comment in the prototypical than in the non-prototypical aggressor condition (*Hypothesis 1*). Moreover, we hypothesized that participants would regard the aggressive behavior as more normative in the prototypical than in the non-prototypical aggressor condition (*Hypothesis 2*). Finally, we predicted that the effect of prototypicality on participants' aggressive behavior would be mediated by their perceptions of normativity (*Hypothesis 3*). Since gender may play a role in the intention to join an aggressor (Bastiaensens et al., 2014), responses of male and female participants were compared. The research was approved by the Ethics Committee at the first author's institution. Written informed consent was obtained from each participant (both in the pilot and in the main study).

## 2 | PILOT STUDY

### 2.1 | Methods

#### 2.1.1 | Participants

Forty-four university students (22 women; mean age = 24.57,  $SD = 6.65$ ) participated in the study on a voluntary basis.


#### 2.1.2 | Procedure and measures

A pilot study was conducted to assess the prevalent belief of university students about immigrants' integration and to ascertain which of the three fictitious users who took part in the conversation would be rated as the most prototypical. Participants were approached by the experimenter in different university buildings and invited to take part in the study, which was presented as part of a broader research on communication exchange on social media. Participants individually completed an online survey in a laboratory. They were first asked to indicate their opinion about immigrants' integration in Italy (1 = *totally disagree*; 7 = *totally agree*), to make salient their pro-immigration position (generally shared among university students). Next, they were presented with an article excerpt concerning the possible benefits of immigrants' integration, which was followed by a comment section including a conversation between three fictitious social media users (see Table 1, comments 1–12). Each user was associated with a default profile icon and a fictitious username, composed of the first name and the surname initial. Six common Italian first names were used, three for female (Michela, Francesca, and Alessandra) and three for male users (Marco, Francesco, and Alessandro). Gender of fictitious users and gender of participants were matched, so that female participants were presented with female user profiles and male participants were presented with male user profiles.

TABLE 1 Fictitious conversation among the four social media users

User	Comment
 Michela F. (Marco F.)	1. Good article! I think that anti-immigration policies make little sense. Even though the need for immigration regulation isn't questionable, we can't turn our back to people who are suffering every day and who are escaping from inhuman living conditions.
 Francesca L.→Michela F. (Francesco L.→Marco F.)	2. I agree. I would also say that media often tend to exaggerate the differences between Europeans and people from the countries listed in the article. Of course, there are cultural differences, but we cannot forget million years of evolution, which make a person born in Syria just the same as a person who was born in Europe.
 Michela F.→Francesca L. (Marco F.→Francesco L.)	3. However, cultural differences are an important issue and may lead to tension... It must be said that they could be a source of enrichment if there are integration programs!
 Alessandra M. (Alessandro M.)	4. I often think about those people who are convinced that immigrants are stealing our jobs. If a foreigner who doesn't speak your language, without a high-level education and comes from a country destroyed by the war is able to be hired instead of you, then you should go over your curriculum vitae.
 Francesca L. (Francesco L.)	5. Unfortunately, migrants have become the scapegoat to justify many problems of the European countries.
 Michela F.→Francesca L. (Marco F.→Francesco L.)	6. However, studies conducted by the CReAM demonstrate that in countries like England, migrants, once integrated into the labor market, have paid more taxes compared to the financial aid they received by the government.
 Alessandra M.→Francesca L. (Alessandra M.→Francesco L.)	7. Migration phenomena have always characterized human history and it makes no sense to try to stop them. With the correct integration policies, there could be growth for everyone. A country that relies exclusively on its own resources and citizens cannot simply exist.
 Francesca L.→Alessandra M. (Francesco L.→Alessandro M.)	8. I agree with the fair/right integration policy statement. Europe should invest in this, and not just clumsily settle migrants in an attempt to solve short-term problems.
 Michela F. (Marco F.)	9. Irrespective of political ideology, Europe has a duty to help asylum-seekers coming from war zones. We need to act starting from the refugee camps. At Idomeni, for example, migrants don't have adequate shelters nor sufficient humanitarian aid. The authorities regularly use violence. There are thousands of people fleeing from suffering and, if they don't die during the journey, they will find it at EU borders. No human being should live like this.
 Francesca L.→Michela F. (Francesco L.→Marco F.)	10. Not to mention the problem regarding human traffickers, who take advantage of the suffering of others. This, unfortunately, is a matter of difficult resolution.
 Alessandra M.→Michela F. (Alessandro M.→Marco F.)	11. And how can we not remember the fact that countries such as Europe or America have a thousand-year-old immigration history!
 Alessandra M. (Alessandro M.)	12. There are many areas that benefit from immigration, and many jobs are based on the interaction between people of different cultures. Just think about science and academic researchers.
 Elisa T. (Luca T.)	13. I think you are all deluded. Tell me how such different cultures could possibly integrate with European society. What is more, Europe is not able to handle immigration and all the rumors about these presumed economic benefits is a self-righteousness fiction. At some point, ethics and moral beliefs must be set aside. Immigration is ruining Europe, so it has to be stopped.

TABLE 1 (Continued)

User	Comment
 Michela F./Francesca L.→Elisa T. (Marco F./ Francesco L.→Luca T.)	14. Can't you see you're an idiot? Being against immigration doesn't make any sense, since it is a normal fact thanks to which countries and cultures were always born. Like all the others of your specie, you don't have full knowledge of the facts to make these statements. You're the usual know-it-all judge who comes here and necessarily says whatever you think and our poor eyes must see this. Do a favor to the users of this website and go post somewhere else.

Note. Comments 13 and 14 were used only in the main study; the user posting comment 14 was Michela/Marco in the prototypical aggressor condition and Francesca/Francesco in the non-prototypical aggressor condition.

The conversation consisted of 12 comments (four for each user), all of which expressed a positive opinion toward immigrants' integration. The stimulus materials were appositely developed to resemble a standard social media webpage. Profile names were colored in Facebook's blue (RGB code 59,89,151) and some comments were indented to represent replies to other comments; a scroll bar allowed to scroll the screen up and down. On the next page, participants were asked to indicate which of the three users best represented the position expressed in the article excerpt (perceived prototypicality of the ingroup, i.e., those who favor immigrants' integration), to what extent they liked the comments posted by each of the three users (evaluation; 1 = *totally dislike*; 7 = *totally like*), and their agreement with the position expressed in the article (1 = *totally disagree*; 7 = *totally agree*).

## 2.2 | Results

Pilot study results revealed a favorable opinion about immigrants' integration among university students ( $M = 4.70$ ,  $SD = 1.46$ , 95% confidence interval [CI] = [4.26, 5.15]),  $t(43) = 3.21$ ,  $p = .003$ ,  $d = .48$  (difference from the mid-point of the scale). Moreover, the three fictitious users were perceived to represent the position expressed in the article to a different extent:  $\chi^2(1) = 7.41$ ,  $p = .025$ . Michela/Marco was rated as the most prototypical by 23 participants (*standardized residuals*: 2.16), Alessandra/Alessandro by 12 participants (*standardized residuals*: 0.70), and Francesca/Francesco by 9 participants (*standardized residuals*: 1.49), respectively. Perceived prototypicality was not affected by gender of participants,  $\chi^2(2) = 0.16$ ,  $p = .926$ . The mixed 2(Gender of participants)  $\times$  3(Target: Michela/Marco vs. Alessandra/Alessandro vs. Francesca/Francesco) ANOVA applied to evaluation of users' comments showed that neither the main effects of target ( $M_{\text{Michela/Marco}} = 4.84$ ,  $SD = 1.55$ , 95% CI = [4.37, 5.31];  $M_{\text{Alessandra/Alessandro}} = 4.59$ ,  $SD = 1.52$ , 95% CI = [4.13, 5.05];  $M_{\text{Francesca/Francesco}} = 4.64$ ,  $SD = 1.46$ , 95% CI = [4.19, 5.08]),  $F(2,41) = 1.21$ ,  $p = .309$ , and gender of participants,  $F(2,42) = 0.63$ ,  $p = .803$ , nor their interaction were significant,  $F(2,41) = 3.01$ ,  $p = .067$ . Finally, participants reported a substantial agreement with the position expressed in the article ( $M = 4.93$ ,  $SD = 1.81$ , 95% CI = [4.38, 5.48]),  $t(43) = 3.42$ ,  $p = .001$ ,  $d = .52$  (difference from the mid-point of the scale). Overall, replicating previous research (e.g., Boccato et al., 2015;

Vezzali et al., 2015), the pilot study indicated that university students' prevalent attitude toward immigrants' integration was positive and that Michela/Marco was perceived as the most prototypical, even though the three users were liked to a similar extent.

## 3 | MAIN STUDY

### 3.1 | Methods

#### 3.1.1 | Participants

Participants were 100 university students (50 women; mean age = 23.91,  $SD = 2.92$ ) who participated in the study on a voluntary basis.

#### 3.1.2 | Procedure

As in the pilot study, participants were approached by the experimenter in different university buildings and invited to take part in the study. After signing informed consent, they were asked to report their opinion about immigrants' integration and were subsequently exposed to the same blog entry used in the pilot study. They read the article excerpt and the fictitious conversation between the three users. Prototypicality of the aggressor was manipulated by introducing at the end of the group discussion a provocative comment by a fourth fictitious user (Elisa/Luca) and an aggressive reply by one of the three initial users (see Table 1). The provocative comment explicitly opposed immigrants' integration. In line with SCT, this should have enhanced the salience of social categorization (i.e., the distinction between those who favor and those who oppose integration of immigrants; pre-test results showed that students fall into the first category) and increased the perceived prototypicality of the ingroup member who expressed support for integration to a greater extent. In line with the results found in the pre-test, the aggressive comment was posted by Michela/Marco in the prototypical aggressor condition and by Francesca/Francesco in the non-prototypical aggressor condition. The words "prototypical" and "non-prototypical" should not be intended in absolute terms but rather in relative terms. It is worth noting that individuals are sensitive to differences in prototypicality (Hogg, 2001) when an ingroup category is made salient. Therefore,

even though one may reasonably think that a relatively moderate difference in prototypicality (as the one between Michela/Marco and Francesca/Francesco) will result in a feeble manipulation, it can be preferable compared to a more extreme difference between a prototypical and a non-prototypical member, because in the latter case the non-prototypical member would likely be perceived as a deviant. Participants were randomly assigned to one of the two experimental conditions.

### 3.1.3 | Measures

The two dependent variables were assessed by asking participants to indicate perceived normativity of the aggressive comment ("To what extent do you think the comment of [name of the prototypical/non-prototypical aggressor] is acceptable?; 1 = *not at all*; 7 = *completely*) and to post a comment as to continue the conversation ("Please post a comment in response to the user Elisa/Luca as if you intended to continue the conversation. It is important for us that you express yourself freely using the words, expressions, and tones that best represent your thoughts. Remember that your comment will always remain anonymous;" aggressive/nonaggressive comment). Participants' comments were coded as aggressive or nonaggressive by two independent coders; discrepancies were solved by a third independent coder. All coders were blind to conditions and unaware of study hypotheses.

## 3.2 | Results

Participants reported a favorable opinion about immigrants' integration ( $M = 5.04$ ,  $SD = 1.72$ ),  $t(99) = 6.04$ ,  $p < .001$ ,  $d = .61$  (difference from the mid-point of the scale). The 2(Experimental condition: prototypical vs. non-prototypical aggressor)  $\times$  2(Gender of participants: female vs. male) ANOVA applied to this item revealed no significant effect,  $F_s(1,96) \leq 1.94$ ,  $p \geq .17$ ,  $\eta_p^2 \leq .02$ .

Coders showed a satisfactory degree of accordance relative to the coding of aggressive/nonaggressive comments (Cohen's kappa = .77). An example of aggressive comment was: "The fact that you are convinced that immigration must be stopped indicates that you are the usual ignorant who mixes the concept of immigration in general with the phenomenon of people fleeing wars. How can you be against immigration, genius?"; an example of nonaggressive comment was: "I'm not afraid of immigration but of people who already live in Europe, those who lash out at the weakest and find excuses to do so". The distribution of aggressive/nonaggressive comments was examined as a function of experimental condition and gender of participants (see Table 2). The chi-square test revealed that aggressive posts were differentially distributed in the two experimental conditions:  $\chi^2(1) = 28.57$ ,  $p < .001$ ,  $\phi = .53$ , but similarly distributed between female and male participants:  $\chi^2(1) = 1.79$ ,  $p = .18$ ,  $\phi = .13$ . As indicated by standardized residuals (see Table 2), the observed frequency of aggressive comments was significantly higher than the expected value in the prototypical aggressor condition and significantly lower than the expected value in the non-prototypical aggressor condition. These results support *Hypothesis 1*.

The 2 (Experimental condition)  $\times$  2 (Gender of participants) ANOVA applied to perceived normativity of the member's comment revealed a significant main effect of experimental condition,  $F(1,96) = 37.21$ ,  $p < .001$ ,  $\eta_p^2 = .14$ , with higher perceived normativity reported in the prototypical ( $M = 4.12$ ,  $SD = 1.57$ , 95% CI = [3.67, 4.57]) than in the non-prototypical aggressor condition ( $M = 2.90$ ,  $SD = 1.63$ , 95% CI = [2.44, 3.66]). No other significant effect was found,  $F_s(1,96) \leq 3.42$ ,  $p_s \geq .067$ ,  $\eta_p^2 \leq .034$ . These results provide support to *Hypothesis 2*.

To test the hypothesized mediation effect, we used the PRO-CESS macro (version 2; Hayes, 2012). Model 4 was tested with 1000 bootstrap resamples, entering experimental condition as the independent variable, aggressive/nonaggressive comments as the dependent variable, and perceived normativity as the mediator. The total effect of experimental condition on aggressive/nonaggressive comments was significant ( $b = 0.47$ ,  $SE = 0.79$ ,  $z = 3.78$ ,  $p = .0002$ ),

	Participants' comment			
	Aggressive		Nonaggressive	
	Observed frequency	Standardized residuals	Observed frequency	Standardized residuals
<i>Experimental condition</i>				
Prototypical aggressor	26	3.2	24	-2.0
Non-prototypical aggressor	2	-3.2	48	2.0
<i>Gender of participants</i>				
Females	11	-0.8	39	0.5
Males	17	0.8	33	-0.5

**TABLE 2** Likelihood of aggressive posts as a function of experimental condition and gender of participants

thus replicating the result obtained with the chi-square test (aggressive comments more frequently observed in the prototypical aggressor condition). The indirect effect via perceived normativity was also significant: estimate = 0.47, 95% bootstrap LLCI = 0.07, 95% bootstrap ULCI = 1.20. This result provides support for *Hypothesis 3*, suggesting that prototypicality enhances the likelihood of aggressive behaviors via perceptions of normativity.

With the aim of exploring potential interaction effects of personal characteristics, we tested two additional models, including gender and opinion towards immigrants' integration as moderators (moderated mediation models; PROCESS version 2, model 5), respectively. Results showed that the interaction gender  $\times$  prototypicality was nonsignificant ( $b = -0.71$ ,  $SE = 1.58$ ,  $z = -0.45$ ,  $p = .65$ ). The overall indirect effect via perceived normativity was significant (estimate = 0.40, 95% bootstrap LLCI = 0.007, 95% bootstrap ULCI = 1.28). Similarly, the interaction opinion  $\times$  prototypicality was nonsignificant:  $b = -0.98$ ,  $SE = 0.94$ ,  $z = 1.03$ ,  $p = .30$ . The overall indirect effect via perceived normativity was significant: estimate = 0.47, 95% bootstrap LLCI = .04, 95% bootstrap ULCI = 1.34. Thus, neither gender nor opinion towards immigrants' integration moderated the effect of prototypicality.

## 4 | DISCUSSION

In this study, we tested prototypicality of the aggressor as a group-level situational factor explaining the spread of aggression on social media. Prototypicality refers to the perceived similarity of a group exemplar to the group prototype (i.e., the defining attributes of the group; Turner et al., 1987). SCT predicts greater conformity to the attitudes, beliefs, and behaviors of highly prototypical ingroup members. This is expected to occur because representations of the self and others are depersonalized, namely defined in accordance with group stereotypical dimensions. Based on these considerations, we predicted that aggression on social media would spread to bystanders to a greater extent when the aggressor was more rather than less prototypical. Moreover, we expected this effect to be mediated by a shift in participants' perceptions of normativity of the aggressive behavior. Results supported our hypotheses.

These findings are in line with SCT and extend it by showing for the first time that prototypical exemplars influence the acceptability and enactment of collective aggressive behaviors in online interactions. The findings of the present study are also in agreement with the social identity model of deindividuation effects (SIDE; Postmes et al., 1998; Reicher et al., 1995), which postulates that in computer-mediated communication (CMC) social influence and adherence to social norms (and, therefore, to the group prototype) might be accentuated under a shared group identity. Indeed, our results suggest that participants may have shifted their self-concept toward a shared social identity (i.e., "those in favor of immigrants' integration"), also fostered by the provocative comment of a fourth user that enhanced salience of categorical distinctions between pro-integration and

anti-integration positions. As a result, they conformed to the behavior of the prototypical aggressor.

Our findings have important implications for the understanding and prevention of online aggression, specifically of forms of aggression such as cyberbullying where bystanders can take an active role by either aiding the victim or taking the side of the aggressor. Our results indicate that if the aggressor is highly prototypical of an ingroup, bystanders are more likely to aggress the victim, even in the absence of an explicit group norm suggesting the acceptability of aggression. However, aggression is unlikely to occur when the aggressor is less prototypical. Indeed, we suggest that prototypicality is particularly likely to play a role in those online group interactions where group norms about cyberaggression are not explicit or well defined, such as online discussion forums or other social network interactions that occur on an occasional basis. In all these contexts, prototypicality acts as a signal indicating the behaviors that are appropriate, thus contributing to form the norm. Future research should investigate the role of prototypicality in contexts with initial explicit (positive or negative) group norms about cyberaggression and test whether prototypical exemplars may alter the existing norms.

Overall, our results provide support for the contention that group-level situational factors should be a primary focus in the study of group processes in online communities. Although we acknowledge the importance of personal factors as determinants of behavior, it is worth noting that group-level factors are still too often overlooked in the analysis of group processes, as is the study of the interaction between contextual/situational (e.g., group-level) and individual factors. In line with this consideration, in the present study, we tested the moderating role of gender and the personal opinion of participants. The fact that these two factors did not interact with prototypicality provides further support for the importance of group-level factors in the context of online aggression. However, it is worthwhile noting that the nonsignificant interactions observed in the present study might depend also on the nature of the variables considered. As to gender, contrary to findings of previous research (Bastiaensens et al., 2014), which underlined gender differences in online aggression, in the present study female and male participants were equally likely to behave aggressively and their normativity perceptions were not significantly different. One possible explanation for this finding is that the experimental setting might have obscured a categorization of the self in terms of gender, both because of the topic of discussion and of the gender matching adopted in the experimental design (the fictitious users were presented as female users to female participants and as male users to male participants). This suggests that gender differences found in previous studies might be contingent on the activation of a categorization of the self in terms of one's gender. Future studies should investigate this possibility. Concerning the opinion about immigrants' integration, as we have already noted, university students typically hold positive attitudes on this issue. Therefore, this variable is likely to reflect, at least in part, a shared social identity. Even though neither of the two interactions tested in the present study was significant, we believe that studying the

interplay of contextual and individual factors in the context of cyberbullying behaviors is a major task for future research.

It is worth noting some limitations of the present experiment. Unlike real online discussion forums, our participants did not spontaneously choose to read the article and the conversation, in accordance with their own personal preferences. However, we suggest that the effects observed in this study could be even stronger outside the laboratory, as people who deliberately choose to join a conversation are likely to identify with the topic discussed, with the result of greater salience of group membership. A second limitation is that participants might have felt pushed to post a comment by the experimental setting, even if they were given the option of not answering. However, this does not affect the validity of results, as in both experimental conditions they were free to post either aggressive or nonaggressive comments. A third limitation is that our sample included university students (i.e., individuals with high-level education) and this limits the generalizability of our results. The findings of the present study should be replicated by using different samples and different topics of discussion. A fourth limitation is that we did not explicitly introduce a social category (i.e., people favoring immigrants' integration) nor asked participants whether they identified with this category. This was done to maintain the conditions of online communities, where often there is no explicit social category of reference (e.g., online discussion forums). However, we assessed the position of participants, that in line with previous research declared to be favorable to integration of immigrants. Furthermore, in accordance with the meta-contrast principle, reading a provocative comment that is opposed to the prevalent opinion expressed in the conversation should have enhanced the salience of a shared social category and increased perceived differences in prototypicality of group members. A further limitation concerns the concurrent assessment of perceived normativity and aggressive comments which does not allow to draw inferences about the causal relationship between these variables. Based on SCT, we hypothesized that prototypicality would affect perceived normativity of the aggressor's behavior and this, in turn, would enhance the likelihood of posting an aggressive comment. Future studies should provide a more solid test of this hypothesis, by examining the effects of experimentally manipulated perceptions of normativity on participants' responses.

In sum, the present experiment provides for the first time evidence that collective participation in aggression on social media is influenced by prototypicality of the aggressor. Furthermore, the findings of this experiment highlight the need for researchers to further consider the role of group-level situational factors in the study of collective forms of online aggression. These findings may ultimately inform social media experts on the psychological underpinnings of aggression perpetration and might help to further understand why and how specific types of social media may be more prone to foster negative relational dynamics.

## CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Open Science Framework at <https://osf.io/396dg/>, reference number DOI: 10.17605/OSF.IO/396DG. The data that support the findings of this study are available on request from the corresponding author.

## ORCID

Elena Trifiletti  <https://orcid.org/0000-0002-9203-507X>

Loris Vezzali  <https://orcid.org/0000-0001-7536-9994>

Soraya E. Shamloo  <http://orcid.org/0000-0003-3640-5504>

Martina Faccini  <https://orcid.org/0000-0003-3105-6886>

Veronica M. Cocco  <https://orcid.org/0000-0002-1777-397X>

## REFERENCES

- Abrams, D., Wetherell, M., Cochran, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology*, 29(2), 97–119. <https://doi.org/10.1111/j.2044-8309.1990.tb00892.x>
- Alonzo, M., & Aiken, M. (2004). Flaming in electronic communication. *Decision Support Systems*, 36, 205–213. [https://doi.org/10.1016/S0167-9236\(02\)00190-2](https://doi.org/10.1016/S0167-9236(02)00190-2)
- Aoyama, I., Barnard-Brak, L., & Talbert, T. (2011). Cyberbullying among high school students: Cluster analysis, sex and age differences and the level of parental monitoring. *International Journal of Cyber Behavior, Psychology and Learning*, 1, 25–35. <https://doi.org/10.4018/ijcbpl.2011010103>
- Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature Communications*, 8(1), 1–8. <https://doi.org/10.1038/ncomms14753>
- Barlińska, J., Szuster, A., & Winiewski, M. (2013). Cyberbullying among adolescent bystanders: Role of the communication medium, form of violence, and empathy. *Journal of Community & Applied Social Psychology*, 23, 37–51. <https://doi.org/10.1002/casp.2137>
- 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>
- Bastiaensens, S., Vandebosch, H., Poels, K., Van Cleemput, K., DeSmet, A., & De Bourdeaudhuij, I. (2014). Cyberbullying on social network sites: An experimental study into bystanders' behavioural intentions to help the victim or reinforce the bully. *Computers in Human Behavior*, 31, 259–271. <https://doi.org/10.1016/j.chb.2013.10.036>
- Boccatto, G., Capozza, C., Trifiletti, E., & Di Bernardo, G. A. (2015). Attachment security and intergroup contact. *Journal of Applied Social Psychology*, 45, 629–647. <https://doi.org/10.1111/jasp.12325>
- Bond, R. M., & Bushman, B. J. (2017). The contagious spread of violence among us adolescents through social networks. *American Journal of Public Health*, 107, 288–294. <https://doi.org/10.2105/AJPH.2016.303550>
- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, 97–102. <https://doi.org/10.1016/j.paid.2014.01.016>
- Cacioppo, J. T., Fowler, J. H., & Christakis, N. A. (2009). Alone in the crowd: The structure and spread of loneliness in a large social network. *Journal of Personality and Social Psychology*, 97(6), 977–991. <https://doi.org/10.1037/a0016076>
- Calvete, E., Orue, I., Estévez, A., Villardón, L., & Padilla, P. (2010). Cyberbullying in adolescents: Modalities and aggressors' profile.



- Computers in Human Behavior*, 26, 1128–1135. <https://doi.org/10.1016/j.chb.2010.03.017>
- Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4), 370–379. <https://doi.org/10.1056/NEJMsa066082>
- Christakis, N. A., & Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358(21), 2249–2258. <https://doi.org/10.1056/NEJMsa0706154>
- Christakis, N. A., & Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks. *PLOS One*, 5(9), e12948. <https://doi.org/10.1371/journal.pone.0012948>
- Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine*, 32(4), 556–577. <https://doi.org/10.1002/sim.5408>
- Fagan, J., Wilkinson, D. L., & Davies, G. (2007). Social contagion of violence. In D. Flannery, A. Vazsonyi, & I. Waldman (Eds.), *The Cambridge handbook of violent behavior* (pp. 688–723). Cambridge University Press.
- Gravani, M., Soureti, A., & Stathi, S. (2018). Using nostalgia to reduce prejudice toward immigrants. *European Journal of Social Psychology*, 48(2), O168–O174. <https://doi.org/10.1002/ejsp.2294>
- Green, B., Horel, T., & Papachristos, A. V. (2017). Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014. *JAMA Internal Medicine*, 177(3), 326–333. <https://doi.org/10.1001/jamainternmed.2016.8245>
- Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling. *White paper*. <http://www.afhayes.com/public/process2012.pdf>
- Hogg, M. A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review*, 5(3), 184–200. [https://doi.org/10.1207/S15327957PSPR0503\\_1](https://doi.org/10.1207/S15327957PSPR0503_1)
- Jordan, J. J., Rand, D. G., Arbesman, S., Fowler, J. H., & Christakis, N. A. (2013). Contagion of cooperation in static and fluid social networks. *PLOS One*, 8(6), e66199. <https://doi.org/10.1371/journal.pone.0066199>
- Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4), 1073–1137.
- Kwon, K. H., & Gruzd, A. (2017). *Is aggression contagious online? A case of swearing on Donald Trump's campaign videos on Youtube*. Proceedings of the 50th Hawaii International Conference on System Sciences. <https://doi.org/10.24251/HICSS.2017.262>
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, 26, 1536–1546. <https://doi.org/10.1016/j.chb.2010.05.023>
- Moore, M. J., Nakano, T., Enomoto, A., & Suda, T. (2012). Anonymity and roles associated with aggressive posts in an online forum. *Computers in Human Behavior*, 28, 861–867. <https://doi.org/10.1016/j.chb.2011.12.005>
- Pfeffer, J., Zorbach, T., & Carley, K. M. (2014). Understanding online firestorms: Negative word-of-mouth dynamics in social media networks. *Journal of Marketing Communications*, 20(1–2), 117–128. <https://doi.org/10.1080/13527266.2013.797778>
- Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? Side effects of computer-mediated communication. *Communication Research*, 25, 689–715. <https://doi.org/10.1177/009365098025006006>
- Powell, K., Wilcox, J., Clonan, A., Bissell, P., Preston, L., Peacock, M., & Holdsworth, M. (2015). The role of social networks in the development of overweight and obesity among adults: A scoping review. *BMC Public Health*, 15(1), 1–13. <https://doi.org/10.1186/s12889-015-2314-0>
- Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology* (Vol. 6, pp. 161–198). Wiley.
- Rosenquist, J. N., Fowler, J. H., & Christakis, N. A. (2011). Social network determinants of depression. *Molecular Psychiatry*, 16(3), 273–281. <https://doi.org/10.1038/mp.2010.13>
- Rösner, L., & Krämer, N. C. (2016). Verbal venting in the social web: Effects of anonymity and group norms on aggressive language use in online comments. *Social Media & Society*, 2(3), 371. <https://doi.org/10.1177/2056305116664220>
- Rost, K., Stahel, L., & Frey, B. S. (2016). Digital social norm enforcement: Online firestorms in social media. *PLOS One*, 11(6), e0155923. <https://doi.org/10.1371/journal.pone.0155923>
- Runions, K. C., & Bak, M. (2015). Online moral disengagement, cyberbullying, and cyber-aggression. *Cyberpsychology, Behavior and Social Networking*, 18(7), 400–405. <https://doi.org/10.1089/cyber.2014.0670>
- Tajfel, H. (1978). *Differentiation between social groups: Studies in the Social Psychology of intergroup relations*. Academic Press.
- Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behaviour. In E. J. Lawler (Ed.), *Advances in group process: Theory and research* (Vol. 2, pp. 77–122). JAI Press.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Blackwell.
- Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, 25, 237–252. <https://doi.org/10.1111/j.2044-8309.1986.tb00732.x>
- Van Cleemput, K., Vandebosch, H., & Pabian, S. (2014). Personal characteristics and contextual factors that determine “helping,” “joining in,” and “doing nothing” when witnessing cyberbullying. *Aggressive Behavior*, 40, 383–396. <https://doi.org/10.1002/ab.21534>
- Vandebosch, H., & Van Cleemput, K. (2009). Cyberbullying among youngsters: Profiles of bullies and victims. *New Media & Society*, 11, 1349–1371. <https://doi.org/10.1177/1461444809341263>
- Vezzali, L., & Giovannini, D. (2010). Social dominance orientation, realistic and symbolic threat: Effects on Italians' acculturation orientations, intergroup attitudes and emotions toward immigrants. *Testing, Psychometrics, Methodology in Applied Psychology*, 17(3), 141–159.
- Vezzali, L., Stathi, S., Giovannini, D., Capozza, D., & Trifiletti, E. (2015). The greatest magic of Harry Potter: Reducing prejudice. *Journal of Applied Social Psychology*, 45, 105–121. <https://doi.org/10.1111/jasp.12279>
- Wachs, S. (2012). Moral disengagement and emotional and social difficulties in bullying and cyberbullying: Differences by participant role. *Emotional and Behavioural Difficulties*, 17(3–4), 347–360. <https://doi.org/10.1080/13632752.2012.704318>
- Yokotani, K., & Takano, M. (2021). Social contagion of cyberbullying via online perpetrator and victim networks. *Computers in Human Behavior*, 119, 106719. <https://doi.org/10.1016/j.chb.2021.106719>
- Zych, I., Baldry, A. C., Farrington, D. P., & Llorent, V. J. (2019). Are children involved in cyberbullying low on empathy? A systematic review and meta-analysis of research on empathy versus different cyberbullying roles. *Aggression and Violent Behavior*, 45, 83–97. <https://doi.org/10.1016/j.avb.2018.03.004>

**How to cite this article:** Trifiletti, E., Giannini, M., Vezzali, L., Shamloo, S. E., Faccini, M., & Cocco, V. M. (2022). At the core of cyberaggression: A group-based explanation. *Aggressive Behavior*, 48, 85–93. <https://doi.org/10.1002/ab.22001>