Bystander Intervention in Cyberbullying and Online Harassment: The Role of Expectancy Violations

NICHOLAS BRODY
University of Puget Sound, USA

The present study extends expectancy violations theory to investigate bystander behavior during cyberbullying and online harassment incidents. Adult participants (N = 368) recalled a recent incident of cyberbullying or online harassment that they observed as a witness and assessed the expectedness and valence of the incidents. Results suggest that incidents were largely seen as negative and as violations of the participants’ expectations. Furthermore, the negative valence of the incident moderated the effects of unexpectedness on bystander behavior. Specifically, the more participants saw incidents as unexpected and negative, the less likely they were to passively stand by and observe the incident or join in with the perpetrators. Theoretical and practical implications and limitations of the findings are discussed.

Keywords: cyberbullying, online harassment, bystander intervention, expectancy violations theory

Extant research has now documented the occurrence of cyberbullying and online harassment. Although estimates vary widely because of definitional ambiguity and methodological differences, Selkie, Fales, and Moreno (2016) conducted a systematic review of more than 1,000 studies examining the prevalence of cyberbullying and online harassment in U.S. adolescents. They found that between 3% and 72% of teens reported being victimized at some point, depending on the time range being examined (i.e., over the past month, year, or ever). Among college students, 19% of participants in one study had been targeted by cyberbullying on a social networking site, and almost half (46%) had witnessed an incident of cyberbullying (Gahagan, Vaterlaus, & Frost, 2016). A nationwide survey of online adults found that 66% of participants had witnessed harassment online, and 41% had been targeted directly (Duggan, 2017). Victims experience a wide array of negative outcomes, including anxiety, depression, stress, and suicide ideation (Kowalski, Giumetti, Schroeder, & Lattanner, 2014). Clearly, cyberbullying and online harassment are relatively common behaviors that Internet users are exposed to directly—as victims or perpetrators—or indirectly as bystanders.

Several studies (e.g., Brody & Vangelisti, 2016; DeSmet et al., 2013) have explored bystander behavior in cyberbullying incidents. Much of that research has invoked the bystander intervention model (BIM; Latané & Darley, 1970) or specific stages of the model. The first two stages of the model argue that,
before assisting during a critical situation, bystanders must notice that the event is occurring and interpret the event as an emergency. The present study explores these first two stages in an online context by synthesizing the first two steps of the BIM with expectancy violations theory (EVT; Burgoon & Hale, 1988). Because of the relational nature of cyberbullying and online harassment (Brody & Vangelisti, 2017), an interpersonal theory such as EVT can supplement the BIM to provide insight into how individuals react to incidents as bystanders when norms are ambiguous. In addition, EVT proposes expectancies as a concept that, when integrated with the BIM, could trigger specific responses in the first two stages of the BIM. Thus, this research explores how norms and expectations in online contexts might lead bystanders to notice a cyberbullying or online harassment incident, interpret it as an emergency, and subsequently take action.

**Cyberbullying and Online Harassment**

Researchers and practitioners have struggled to agree on a shared definition of cyberbullying (Englander, Donnerstein, Kowalski, Lin, & Parti, 2017). Key questions include how cyberbullying differs from traditional bullying, and how cyberbullying differs from other antisocial online behaviors such as online harassment and cyberstalking. Traditional bullying is defined as intentional, repeated behaviors that harm others (Olweus, 1999). Bullying incidents often reveal a power imbalance between perpetrator and victim (Olweus, 1999). Cyberbullying has been defined similarly as the use of technology to harm others, but definitions often exclude the repetition and power imbalance criteria because of the nature of communication in online environments. For instance, one incident of harm in an online context can have repeated effects as it is reshared (Kazerooi, Taylor, Bazarova, & Whitlock, 2018), and power imbalance is sometimes contingent on anonymity (Smith, del Barrio, & Tokunaga, 2013). Shultz, Heilman, and Hart (2014) note that most previous definitions rely on bystanders’ judgments to determine whether an incident is severe enough to be categorized as cyberbullying.

Online harassment is a broader range of behaviors that includes the use of technology to harm others (Wolak, Mitchell, & Finkelhor, 2007). Most important for the present study, bystanders to such incidents cannot usually distinguish between what constitutes cyberbullying versus online harassment or other forms of harmful online behaviors (Wolak et al., 2007). Thus, because the present study explores the phenomenon from the perspective of the bystanders, I examine both online harassment and cyberbullying incidents. For readability, the term cyberbullying is used in the following review, unless the specified study examined a different behavior.

**Bystander Behavior During Incidents of Cyberbullying and Online Harassment**

Given the semipublic nature of many forms of online communication, cyberbullying incidents often occur in the presence of bystanders. For instance, a text message can be documented with a screenshot and forwarded to dozens or hundreds of other audience members. Social media sites such as Facebook are designed to facilitate the mass sharing of content (boyd, 2014). Thus, an incident of cyberbullying on Facebook might be seen by the entire networks of both the victim and the perpetrator—potentially thousands of individuals.
There are benefits and drawbacks to the visibility of such incidents. When hundreds of bystanders see an incident of cyberbullying, it might exacerbate the pain and embarrassment it causes a victim. In some cases, witnesses can join in on the harassment (O’Connell, Pepler, & Craig, 1999). On the other hand, bystanders have the power to intervene and stop or assuage the incident, either by directly confronting the bully or offering support to the victim (e.g., Brody & Vangelisti, 2016). However, bystanders often decide not to intervene.

In one experiment of helping behavior in response to a series of Facebook bullying scenarios, more than 40% of participants chose not to help at some point in the study (Freis & Gurung, 2013). Dillon and Bushman (2015) conducted an experiment in an online chat context and found that, although 68% of their participants noticed that cyberbullying was occurring, only 10% directly intervened by contacting the bully, victim, or lead researcher. Although most Internet users have witnessed cyberbullying or online harassment, only about a third reported intervening to assist a victim (Duggan, 2017). Overall, there is ample evidence that witnesses often do not intervene to help a victim or confront a perpetrator during online harassment or cyberbullying incidents. There are some elements of online contexts (e.g., anonymity, very large audience sizes) that can complicate the choices made by online bystanders compared with offline bystanders, especially when norms in these contexts are ambiguous (DiFranzo, Taylor, Kazerooni, Wherry, & Bazarova, 2018; Obermaier, Fawzi, & Koch, 2016). For instance, bystanders to cyberbullying are less likely to intervene when they perceive themselves to be visually anonymous (Brody & Vangelisti, 2016).

Researchers have explored what mechanisms might inhibit a witness or bystander from actively intervening by examining the role of bystander intervention in both offline and online environments. This research has invoked a variety of theoretical perspectives, including the BIM (e.g., Latané & Darley, 1970) and social cognitive theory (e.g., Bandura, 1986). Other research has focused on personality and contextual features of the medium in which the bullying is occurring.

### Bystander Behavior From a Social Cognitive Perspective

From a social cognitive perspective, the severity of the bullying incident appears to influence bystander behavior. For instance, participants in a study of secondary school students were more likely to help a victim of bullying on a social networking site when the incident was considered more severe (Bastiaensens et al., 2014). When the participant perceived a close relationship with other bystanders and the incident was severe, participants were more likely to give the victims advice and comfort the victim than when the other bystanders were acquaintances or when the incident was less severe. A social cognitive framework has also been used to examine the role of moral disengagement on bystander behavior (Allison & Bussey, 2017) in middle school students and found that moral disengagement interacted with group norms to predict bystander behavior.

### Bystander Behavior From the Bystander Intervention Model Perspective

Other research has applied the BIM to better understand cyber bystander actions (Darley & Latané, 1968). The BIM describes a five-step process in which witnesses to emergencies might intervene. First, bystanders must (1) notice the event, and then (2) interpret it as an emergency. Subsequently, witnesses
must (3) take personal responsibility for helping, (4) decide on an intervention strategy, and then (5) intervene. Much of the research on bystanders during cyberbullying incidents has examined the final three steps.

Brody and Vangelisti (2016) extended research into the bystander effect (e.g., Darley & Latané, 1968) and found that several factors influenced the propensity of an individual bystander to intervene during a cyberbullying incident on Facebook: the number of other bystanders, perceived anonymity, and the relational closeness to a victim. Bystanders were especially likely to intervene to defend a victim when they perceived themselves to be identifiable (i.e., not anonymous), there were few other bystanders, and when they had a close relationship with the victim. Presumably, each of these factors affected whether the bystanders felt personally responsible for helping. In a direct test of the role of personal responsibility, Obermaier and associates (2016) found that when participants perceived a very low number of witnesses to a Facebook bullying incident (i.e., two compared with 5,025), they felt more personally responsible to intervene, which in turn predicted participants’ intention to intervene. DiFranzo and colleagues (2018) designed a mockup social media site and tracked users’ behavior in response to a cyberbullying incident on the site. Bystanders were more likely to intervene when they perceived a lower audience size, which led them to feel “seen” by other users and thus take personal responsibility for intervening.

Fewer studies into cyberbullying have examined the first two steps of the BIM, but there is some research that suggests a normative approach to investigating bystander intervention. Dillon and Bushman (2015) found that the likelihood of a bystander intervening increases 4.62 times by simply noticing that an incident is occurring. Koehler and Weber (2018) found that cyberbullying incidents that included threats were seen as more severe, which in turn related to a higher willingness to assist victims. When cyberbullying incidents contained insults but no threats, participants did not see the incident as severe and were more likely to engage in victim blaming and less likely to intervene and help a victim (Koehler & Weber, 2018).

Step 2 of the BIM requires that bystanders interpret the event as a situation that warrants helping. In incidents of cyberbullying, this means having at least a working definition of what constitutes bullying. However, as noted above, research into bullying and cyberbullying is plagued by definitional ambiguity (Tokunaga, 2010; Vaillancourt et al., 2008) and surely bystanders to such incidents are similarly unclear as to what might constitute an incident of bullying. Thus, it is likely that norms and expectations of what constitutes acceptable behavior are core aspects of both noticing an event and interpreting the event as one that necessitates helping behavior.

The Role of Norms and Expectancies in Predicting Bystander Behavior

In a classic study of bystander intervention that extended Darley and Latané’s (1968) work on helping behavior in emergency situations, Horowitz (1971) discovered that social norms played a stronger role than number of other bystanders in predicting whether witnesses would help. Members of a service-oriented campus group were more likely to help as number of bystanders increased; members of a social-oriented campus group were less likely to help as the number of bystanders increased. Similarly, Rutkowski, Gruder, and Romer (1983) investigated the effect of group norms on the bystander effect, and found that cohesive groups were less susceptible to the bystander effect than noncohesive groups. Specifically, when a group was cohesive and there was a group norm that tapped the norm of shared social responsibility,
participants were more likely to intervene to help a victim. These findings, although not yet directly replicated in the cyberbullying context, suggest that behavioral and social norms likely exert an influence on bystander behavior in cyberbullying contexts as well.

Norms are generally described as beliefs about what are appropriate behaviors in a particular environment (Aarts & Dijksterhuis, 2003; Cialdini, Reno, & Kallgren, 1990). Beliefs about what behaviors are normative in online environments are important determinants of behavior, but they are often ambiguous. Norms are socially constructed and learned through experience (Preece, 2004), and newer online contexts such as social network sites have less firmly established rules. Moreover, norms are often splintered and distinct for each subgroup of users. Norms can become even more important in online interactions in which participants are anonymous (Postmes, Spears, & Lea, 1998).

Some researchers have framed online behaviors as a consequence of relatively ambiguous online group norms. For instance, O'Sullivan and Flanagin (2003) offered an interactional-normative framework for understanding flaming (i.e., hostile behavior in online environments). Not every online subculture has the same norms, and thus flaming is not seen as overtly negative in every context (O'Sullivan & Flanagin, 2003). Understanding antisocial online behavior is difficult because of the confluence of so many communities, relationships, and individuals online, which leads to multiple conflicting norms in each context. O'Sullivan and Flanagin settle on a definition of flaming as the intentional violation of norms, but leave room for the application of the model to various other forms of mediated harassment (e.g., cyberbullying). One difference between flaming and cyberbullying is the relationship between perpetrator and victim. Whereas flaming often occurs in anonymous online contexts, cyberbullying frequently occurs in situations in which the victim, perpetrator, and witnesses are friends or romantic partners (e.g., Brody & Vangelisti, 2017).

When group identity is not especially salient or relevant, Nicholls and Rice (2017) suggest the use of interpersonal theories such as EVT (Burgoon & Hale, 1988) to explain online behavior when norms are ambiguous. Because of the relational nature of cyberbullying, the present study applies an interpersonal theory of norm violation—EVT—to better understand how norm and rule violation relate to bystander behavior.

**Expectancy Violations Theory and Bystander Behavior**

Norms exert an influence on both sender and receiver behavior, and those norms are often ambiguous. EVT provides an explanatory framework for what happens when perceived norms or rules are violated (Burgoon & Hale, 1988). The EVT framework is especially useful for examining norm violations in online environments when group identity is less salient than individual identity (Nicholls & Rice, 2017). Moreover, there are elements of EVT that allow for the examination of the first two steps of the BIM.

An expectancy violation is "a behavior that a receiver notices as being different from the behavioral display that (s)he expected" (Afifi & Metts, 1998, p. 367). Individuals establish expectations for any situation based on what they expect (or predict) to occur. People derive their expectancies from a number of sources, including the context in which the incident is occurring and their relationship with/the characteristics of their cocommunicator(s) (Burgoon, 1993). In a social networking site context, for instance, an individual may not expect cyberbullying to occur because the site they are using is not anonymous. Or they may not expect
a friend to be cyberbullied because their friend has been self-assured and confident in their previous interactions. When expectancies are violated, bystanders are likely to notice an event, which parallels Step 1 of the BIM.

When predictions are inaccurate (i.e., when individuals’ expectancies are violated), EVT posits that people assign a valence to the violation based on whether the violating behavior exceeds (positive violation) or falls short (negative violation) of what the message receiver expected to happen. When a violation occurs and is evaluated as negative, receivers tend to do less than expected. When the violation is positive, they do more than expected. This outcome is contingent on a number of other variables, including the receiver’s relationship with the violator and level of uncertainty surrounding the event (e.g., Bachmann & Guerrero, 2006). Although initially conceptualized as a way to understand and predict violations of nonverbal norms in interpersonal contexts, the theory has been extended to examine verbal and relational (e.g., Afifi & Metts, 1998; Bachman & Guerrero, 2006) violations, as well as expectancy violations that occur in online contexts (e.g., Bevan, Ang, & Fearn, 2014; McLaughlin & Vitak, 2012; Ramirez & Wang, 2008; Rui & Stefanone, 2018). When the valence of an incident is negative, bystanders might interpret the incident as an emergency that warrants intervention, which parallels Step 2 of the BIM.

Much of this research has examined specific types of online expectancy violations and assessed the communicative and relational implications of such violations. For instance, individuals consider being unfriended on Facebook to be an expectancy violation, and they are more likely to see the unfriending as negatively valenced when enacted by a close friend (Bevan et al., 2014).

McLaughlin and Vitak (2012) conducted focus groups to examine norm violations on Facebook, and framed themes from the qualitative study using EVT. Norms were often implicit and concerned how frequently to post and for which audience, which friends to tag, and how quickly to respond to new friend requests. Participants also noted that their Facebook friends were generally considerate of their self-presentational goals, and did not have issues with friends posting inappropriate content to their pages. That said, one of the more frequently mentioned norm violations occurred when individuals observed “heated interactions, fights, and name calling” (McLaughlin & Vitak, 2012, p. 308) in public or semipublic areas of Facebook. Other research suggests that online arguments and interpersonal disagreements are some of the more frequent types of cyberbullying that occur on Facebook (Brody & Vangelisti, 2017).

Although they did not use an EVT framework, researchers have directly examined the role of norms in predicting cyberbullying behavior. In one study of college-age participants, injunctive norms (rules that emerge from perceived peer approval of a specific behavior) that indicated approval of cyberbullying behavior were positively associated with malicious cyberbullying behavior (i.e., behaviors such as sending rude messages to someone; Doane, Pearson, & Kelley, 2014). Similarly, descriptive norms, which are rules that emerge directly from observation of peer behavior, were positively related to both malicious cyberbullying and public humiliation (i.e., behaviors such as publicly posting an embarrassing picture of someone; Doane et al., 2014). These findings collectively suggest that individuals are more likely to perpetrate cyberbullying behavior if they perceive norms in mediated environments that allow for such behavior.
But how might norms guide the behavior of bystanders? In one investigation of bystander reactions to cyberbullying, Schacter, Greenberg, and Juvonen (2016) investigated whether the depth of an online disclosure predicted bystander behavior. Specifically, it was expected that high-depth disclosures by a victim on Facebook might violate perceived norms regarding sharing too much information online (e.g., “oversharing”), and thus bystanders would assign more blame to the victim and perhaps be less likely to intervene. Indeed, results demonstrated that high-depth (vs. low-depth) disclosures by a cyberbullying target were positively associated with blaming the victim, which in turn negatively related to the likelihood of intervening to assist the victim (Schacter et al., 2016). However, the valence of the disclosure (positive or negative) did not relate to bystander behavior. Schacter and associates explained that individuals might have equally negative reactions to both positive and negative personal disclosures. Positive personal disclosures might be seen as narcissistic, and negative disclosures can be seen as overly needy. These findings underscore the influence of perceived norms on bystander behavior, but the experimental design necessitated a focus on one specific type of norm violation—personal disclosures—that might obscure the role of valence in driving bystander intervention.

Relating to cyberbullying, specifically, DeSmet and colleagues (2014) conducted a series of focus groups with teens to uncover themes relating to how and when they decided to intervene during cyberbullying incidents. Norms and expectations were consistently reported as determinants of the participants’ bystander decisions. Participants thought they were expected to report cyberbullying to their teachers and other school staff members. Others reported norms established by their parents that emphasized not getting involved. Overall, participants expected their friends to help them if they were bullied, but not by directly confronting the perpetrator. Rather, they expected to receive offers of help and support directly from their peers (DeSmet et al., 2014). In another study, participants who read a cyberbullying scenario in which a bystander defended a victim reported higher levels of normative beliefs about helping others compared with participants who read a scenario in which a bystander joined in with a bully (Leung, Wong, & Farver, 2018).

Other research has found that when participants were in groups that reported high levels of moral disengagement, their own moral standards were positively associated with likelihood of intervention during a cyberbullying incident (Allison & Bussey, 2017). However, in classes with low levels of moral disengagement, individual moral disengagement did not relate to bystander behavior.

Overall, previous research suggests that expectancies for proper online behavior vary widely, but online norms play a role in bystander behavior. Although cyberbullying and online harassment have been studied using a normative framework, the present study invokes EVT to examine bystander behavior.

**Applying Expectancy Violation Theory to Predict Bystander Intervention**

In sum, norms are a determining factor in bystanders’ intervention decisions, and EVT (Burgoon & Hale, 1988) provides a framework for predicting such behavior. Individuals have expectancies for how other people should behave that emerge from societal norms about appropriate behavior (Burgoon, 1993). In line with Step 1 of the BIM in which individuals notice an incident, violations of these expectancies arouse the attention of communicators, and can in turn spur them to interpret the meaning of the violation and act
accordingly. Bystanders can take a variety of actions when they notice that cyberbullying has occurred (Salmivalli & Voeten, 2004). They can choose to actively defend the victim by confronting the perpetrator. They might, however, decide not to intervene, instead choosing to passively observe the incident. Or they might assist the perpetrator in harassing the victim.

Based on the expectancy component of EVT, it is likely that the degree to which bystanders feel their expectancies have been violated will be associated with their subsequent behavior. Thus,

\[ H1: \text{The degree to which expectancies are violated is associated with bystander (a) active defending, (b) passive observing, and (c) assisting the perpetrator behavior.} \]

As noted above, however, expectancies rarely affect behavior in isolation. Similar to Step 2 of the BIM in which bystanders interpret an incident as an emergency, Burgoon (1993) posits that, when an expectancy violation occurs, individuals assess the violation and determine whether it is positively valenced or negatively valenced. When violations are interpreted as positive, individuals are subsequently expected to produce positive interaction patterns. However, making specific predictions in the cyberbullying context is not straightforward given that bystanders are simultaneously interpreting the valence of the incident as it relates to the perpetrator and the victim. For instance, a cyberbullying incident might exceed bystander expectations if they think it is funny or inconsequential, but might be seen as quite negative if they perceive harm being done to the victim. Thus, reactions to incidents are predicted to involve an interplay of the degree to which expectations are violated and the valence of the incident:

\[ H2: \text{Violation valence will moderate the effect of the degree of expectancy violation on (a) active defending, (b) passive observing, and (c) assisting the perpetrator behavior.} \]

**Method**

Participants \((N = 368)\) were recruited using the Amazon.com Mechanical Turk website (Mturk) which allows individuals to complete tasks for nominal fees. Sheehan (2018) notes several ethical issues surrounding the usage of Mturk, including power differentials between researchers and workers as well as appropriate payment. To address these concerns, I compensated workers in the present study above the national minimum wage. In addition, responses from Mturk have been found to be reliable (e.g., Buhrmester, Kwang, & Gosling, 2011; Sheehan, 2018). Much previous work into online harassment and bullying has focused on secondary and college students, but as previously noted, more than 40% of Internet-using adults in the United States have experienced online harassment, and more than 60% have witnessed harassment (Duggan, 2017). Thus, Mturk was used to ensure a diverse (age-wise) range of participants. Indeed, the average age of participants in the present study was 34.45 years \((SD = 10.99)\) and ranged from 19 to 79 years.

Of the participants, 164 (45.3%) identified as male and 197 (54.4%) identified as female. One participant did not identify as male or female, and six participants did not report their gender. A majority of participants were Caucasian \((n = 274, 74.4\%)\), followed by African American \((n = 36, 9.7\%)\), Asian/Pacific
Islander (n = 23, 6.3%), Hispanic or Latinx (n = 18, 4.9%), Native American (n = 4, 1%) and other (n = 8, 2.2%). Five participants (1.4%) abstained from reporting their ethnic background.

**Procedure**

Participants were eligible if they had witnessed or observed an incident of harassment or cyberbullying on Facebook in the previous year. Experiences were limited to Facebook to control for possible confounding factors related to the design and usage of other social media sites. In addition, Facebook was (and continues to be) the most frequently used social media site by adults in the United States (Perrin & Anderson, 2019). After confirming their eligibility, participants from Mturk were routed to an online survey where they provided informed consent. Participants were then asked to recall an incident of online harassment or cyberbullying that they had witnessed or observed in the previous year. Cyberbullying and online harassment were defined as when “some

one was targeted on Facebook in an aggressive or threatening manner and they witnessed the event or were aware of the hurtful event soon after it occurred.” Similar to research on hurtful events in relationships (e.g., Vangelisti & Young, 2000), I asked participants to write detailed accounts of what occurred during the incident. They then responded to an open-ended question that asked what happened that led up to the incident. After describing the incident in detail, participants answered several questions (reported in a separate study) about the perceived hurtfulness of the incident, before responding to a third open-ended question about how they responded (if at all) following the incident. Following the open-ended questions, participants completed a variety of closed-ended items, reported below.

**Degree of Expectancy Violation**

The extent to which the incident violated a participant’s expectations was assessed with three semantic differential items from Afifi and Metts (1998). For instance, participants assessed whether the incident “was completely expected” or “not at all expected.” Items were assessed on a 7-point scale (M = 4.93, SD = 1.62, α = .80).

**Valence**

The valence of the incident was measured with three items developed by Johnson (2012). One item was slightly adapted to refer to online behavior instead of language use. Participants assessed the valence of the incident with three Likert-scale items. For instance, participants were asked, “To what extent did the incident meet your expectations about online behavior?” (1 = much more negative than expected, 5 = much more positive than expected). Items were reverse scored so higher scores indicated negative valence (M = 4.05, SD = 0.90, α = .82).

**Bystander Behavior**

Active defending, passive observing, and assisting the perpetrator behaviors were measured with an adapted version of the participant role questionnaire developed by Salmivalli and Voeten (2004). Each subscale contained three items. Because an adult sample was used, wording that referred to school
environments or students was removed or adapted. In addition, the structure of the questions was reworded to refer to a specific incident rather than the participants' role across incidents. One additional subscale—reinforcing—which refers to actively encouraging the perpetrator, did not achieve acceptable reliability ($\alpha = .47$) and was thus removed from analyses. Participants answered a series of Likert-style questions (1 = *strongly disagree*, 7 = *strongly agree*). Active defending (e.g., "I comforted the victim"; $M = 3.97$, $SD = 2.02$, $\alpha = .82$), passive observing (e.g., "I stayed outside the situation"; $M = 3.30$, $SD = 1.80$, $\alpha = .79$), and assisting behavior (e.g., "I joined in with the perpetrator"; $M = 1.57$, $SD = 1.17$, $\alpha = .93$) all demonstrated acceptable reliability. Table 1 displays a correlation matrix of the study variables.

**Table 1. Correlations Among the Variables.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unexpectedness</th>
<th>Negative valence</th>
<th>Active defending</th>
<th>Assisting the perpetrator</th>
<th>Passive observing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexpectedness</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Negative valence</td>
<td>.41***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Active defending</td>
<td>.13*</td>
<td>.15**</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Assisting the perpetrator</td>
<td>—.15**</td>
<td>—.43***</td>
<td>.07</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Passive observing</td>
<td>—.15**</td>
<td>—.15**</td>
<td>—.47***</td>
<td>.20***</td>
<td>—</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.

**Results**

To ensure that the recalled incidents tapped the central variables of EVT (i.e., whether cyberbullying incidents observed by participants constituted an expectancy violation, and whether such violations were perceived as negative), one-sample $t$ tests examined whether participants reported expectancy violation and negative valence levels above the midpoint of the scales. The means and standard deviations for each variable appear in the Method section. Both unexpectedness, $t(365) = 17.022$, $p < .001$, and valence, $t(365) = 33.20$, $p < .001$, were significantly above the midpoints of their respective scales (3.5 for unexpectedness, 2.5 for negative valence). Thus, participants generally saw the cyberbullying incidents they reported as both unexpected and negative.

Hypotheses 1 and 2 examined whether unexpectedness and negative valence related to bystander intervention behavior, and were assessed with a series of hierarchical linear regressions. Each regression investigated a distinct dependent variable (i.e., active defending, passive observing, and assisting). The first block of each regression consisted of the degree of expectancy violation and valence variables, centered to
account for multicollinearity. The second block consisted of the interaction terms for each variable. Table 2 displays the results of the regression.

**Table 2. Hierarchical Regressions Predicting Bystander Behavior.**

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>( B )</th>
<th>( SE , B )</th>
<th>( \beta )</th>
<th>( \Delta R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active defending</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexpectedness</td>
<td>.11</td>
<td>.07</td>
<td>.09</td>
<td>.03**</td>
</tr>
<tr>
<td>Negative valence</td>
<td>.27</td>
<td>.13</td>
<td>.12*</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Unexpectedness × Negative Valence</td>
<td>−.04</td>
<td>.08</td>
<td>−.03</td>
<td></td>
</tr>
<tr>
<td><strong>Passive observing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td>.03**</td>
</tr>
<tr>
<td>Unexpectedness</td>
<td>−.16</td>
<td>.07</td>
<td>−.14*</td>
<td></td>
</tr>
<tr>
<td>Negative valence</td>
<td>−.18</td>
<td>.12</td>
<td>−.09</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td>.01*</td>
</tr>
<tr>
<td>Unexpectedness × Negative Valence</td>
<td>.16</td>
<td>.07</td>
<td>.12*</td>
<td></td>
</tr>
<tr>
<td><strong>Assisting the perpetrator</strong></td>
<td></td>
<td></td>
<td></td>
<td>.18***</td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unexpectedness</td>
<td>−.01</td>
<td>−.01</td>
<td>−.01</td>
<td></td>
</tr>
<tr>
<td>Negative valence</td>
<td>−.56</td>
<td>.07</td>
<td>−.42***</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td>.01*</td>
</tr>
<tr>
<td>Unexpectedness × Negative Valence</td>
<td>.11</td>
<td>.04</td>
<td>.12*</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Active defending: Total  \( R^2 = .03 \), adjusted  \( R^2 = .02 \),  \( F(3, 366) = 3.65, p < .05 \). Passive observing: Total  \( R^2 = .04 \), adjusted  \( R^2 = .04 \),  \( F(3, 361) = 5.49, p < .01 \). Assisting the perpetrator: Total  \( R^2 = .20 \), adjusted  \( R^2 = .19 \),  \( F(3, 363) = 29.74, p < .001 \). *\( p < .05 \). **\( p < .01 \). ***\( p < .001 \).

Participants were more likely to report actively defending the victim when they perceived the incident as negatively valenced. However, the unexpectedness of the incident did not relate to active defending, nor did it interact with the valence to predict active defending. Thus, Hypothesis 1a and Hypothesis 2a were not supported.

In the second regression, which entered passive observing as the criterion variable, there was a negative relationship between unexpectedness and passively observing. However, as predicted, that association was qualified by a significant interaction between valence and unexpectedness. Figure 1 displays a plot of the interaction. The unstandardized simple slope for individuals who reported a high negative valence was −.002 (\( p = .98 \)), the unstandardized simple slope for individuals reporting the mean level of negative valence was −.14 (\( p = .03 \)), and the unstandardized simple slope for individuals reporting a low level of negative valence was −.28 (\( p = .01 \)). For incidents that were perceived as not negatively valenced (1 SD below the mean) or moderately negatively valenced (mean level of negative valence), there were significant, negative relationships between unexpectedness and passive observing. When incidents were perceived as especially negative (1 SD above the mean), there was no relationship between unexpectedness...
and passive observing. Hypothesis 1b and Hypothesis 2b were both supported. Results and implications of the above findings are discussed below.

Figure 1. Interaction between negative valence and unexpectedness predicting passive observing.

Participants were less likely to assist the perpetrator when they perceived the incident as negatively valenced, but that effect was again qualified by a significant interaction between negative valence and unexpectedness. Figure 2 displays a plot of the interaction. The unstandardized simple slope for individuals who reported a high negative valence was −.09 (p = .07), the unstandardized simple slope for individuals reporting the mean level of negative valence was −.01 (p = .81), and the unstandardized simple slope for individuals reporting a low level of negative valence was −.10 (p = .10). For incidents that were perceived as quite negative (1 SD above the mean), the unexpectedness of the incident positively related to assisting the perpetrator. For incidents that were perceived as less negative (1 SD below the mean), the unexpectedness of the incident negatively related to assisting the perpetrator. Hypothesis 2c was supported, but Hypothesis 1c was not supported.
When bystanders perceived a cyberbullying incident to be a violation of their expectations for appropriate behavior, they were less likely to stand by and passively observe the incident, especially when they perceived the incident to be very negative. The results generally provide support for the extension of EVT into the cyberbullying context. The recalled incidents of cyberbullying were seen as both unexpected and negative. Furthermore, the perception of a violation of expectations and negative valence of the incident related to the antisocial bystander behaviors of joining in with the perpetrator and passively observing the incidents. When the valence of the incident was ambiguous or not especially negative, the extent to which the incident violated the bystander’s expectancies was negatively associated with passively observing the incident or joining in with the perpetrator. However, only the negative valence of the incident was associated with the prosocial behavior of actively defending the victim. When individuals perceived an incident as especially negative, they were more likely to step in and assist the victim regardless of whether it violated their expectations.

**Extending Expectancy Violations Theory**

EVT was initially developed as a theory that examined nonverbal space violations (Burgoon & Hale, 1988). In subsequent years, it has been extended to investigate norm violations in a number of interpersonal contexts, such as romantic relationships (Afifi & Metts, 1998). When applied to online contexts, the theory has been applied to unfriending behavior (Bevan et al., 2014) and responses to face-threatening information (Rui & Stefanone, 2018). The results of this study build on this research by both extending EVT into a novel context and providing a new lens for understanding bystander behavior during cyberbullying incidents.
Some specific elements of the theory are especially useful for understanding online bystander behavior. First, results clearly support the notion that cyberbullying incidents were seen as negative expectancy violations. The means for unexpectedness and negative valence were both well above the midpoint of the scale. Thus, study participants likely had a baseline assumption that interactions on Facebook should be kind and free of bullying and harassment. Although the common assumption is that toxic behaviors occur more frequently in an online environment than offline, research has found that negative online behaviors such as flaming are relatively rare as a proportion of the total number of messages exchanged (Derks, Fischer, & Bos, 2008). Although 66% of online adults have witnessed and 41% have personally experienced some form of online harassment in the previous year (Duggan, 2017), these incidents appear to be relatively infrequent, but memorable. In the present study, only 15% of the participants had been personally cyberbullied in the previous three months. Because incidents occur so rarely, when they do occur, they are seen as expectancy violations.

As predicted by EVT, there was an interaction between violation valence and expectedness. When participants’ expectancies were violated, they were less likely to passively stand by and observe the incident, but only when they perceived the valence of the incident as not especially negative. When they perceived the incident to be especially negative, the unexpectedness of the incident did not relate to their passive observing behavior. In other words, the extent to which an incident of cyberbullying was an expectancy violation only had an effect on passive observation when the valence of the incident was ambiguous. When bystanders were not sure how to interpret the severity of the cyberbullying, expectancy violation played a role in their decision making.

The results suggest several implications for the application of EVT into the cyberbullying context. First, cyberbullying and online harassment are generally seen as expectancy violations by witnesses to the instances, which implies that incidents of online harassment and cyberbullying fit within the scope of EVT. However, EVT was only predictive of bystander behavior when the incident was not especially negative. This finding corresponds with Burgoon and Hale’s (1988) initial formulation of EVT, which suggests that more favorable communication outcomes would occur when the expectancy violation is seen as more positively valenced than the expected behavior. However, most other findings did not directly support EVT predictions relating to the interaction between expectancy and violation valence, suggesting that EVT can be used to explain passive observing behavior, but perhaps not more active bystander behaviors.

It is also important to note that active defending and passive observing are distinct behaviors and not perfectly correlated. Thus, bystanders may report that they did not defend a victim (which generally involves directly reaching out to the victim), but also would not characterize their behavior as passively observing. Thus, the model differentially predicted each of these behaviors, which implies that more overt bystander behaviors (e.g., active defending) could perhaps be better explained by other models of bystander behavior, such as the BIM.

**Integrating Expectancy Violations Theory With the Bystander Intervention Model**

The BIM is a five-step model designed to predict bystander behavior in emergency situations (Latané & Darley, 1970). Bystanders must first notice the event and then interpret the event as an
emergency before taking responsibility to help and, ultimately, deciding whether to help or not. The results shed light on what factors influence whether individuals notice a cyberbullying incident or interpret the event as exigent enough to warrant reacting.

Specifically, individuals may have been less likely to passively stand by when the event violated their expectations because the expectancy violation served as a signal that grabbed their attention and prompted them to notice the event (Step 1). This is especially true when the incidents did not seem especially negative. Participants might not have noticed the incident if it had not violated their expectations, which related to their reporting a lower likelihood of passively standing by. Similarly, when incidents were negatively valenced, participants likely interpreted the incidents as emergencies (Step 2), or at least critical situations that warranted a response of some sort. When participants saw an incident as negatively valenced, they were less likely to stand by or join in with the perpetrator regardless of whether the incident violated their expectations. Participants who perceived the incident to be negatively valenced were also more likely to report actively defending the victim. Thus, the extension of EVT in the present study also supports several tenets of the BIM. Specifically, the extent to which online bystanders notice and interpret the event as a norm violation (i.e., a violation of their expectations) negatively affected their propensity to passively stand by and observe the bullying or join in with the perpetrator.

**Practical Implications**

The results offer several practical implications for both bystanders and practitioners/educators who teach individuals about online ethics and decision making. Bystanders were most likely to passively stand by or join in to assist the perpetrator when the incidents they were observing were perceived as expected and not especially negative. However, norms for behavior in online environments are often ambiguous and even the definition of cyberbullying is arbitrary and inconsistent. This might lead to incidents that are harmful and damaging to victims being interpreted by witnesses/bystanders as not especially damaging or noteworthy. Witnesses should be trained to take an “upstander” approach, which encourages witnesses to cyberbullying to take action by default, even when they are not sure whether a particular message is damaging to the recipient.

The findings also have implications for system designers. Other researchers have extended the BIM to suggest that designers integrate audience size indicators to further prompt bystander intervention (e.g., DiFranzo et al., 2018). Designers could also visually flag posts that appear to be cyberbullying or online harassment, which would assist passive viewers to better notice the incident and perhaps decide to intervene.

**Limitations and Future Directions**

The present study investigated two components of EVT: expectancies and violation valence. However, the model also specifies several other moderating variables, including communicator reward valence. The relative cost/benefit of the violator’s relationship with the victim is predicted to influence subsequent behavior. In this investigation, that would mean bystanders would assess their relationship with the perpetrator when deciding whether to intervene. In an online environment where individuals are often connected to people they do not know (and when perpetrators might use fake accounts to maintain
anonymity), it is not always possible to assess this relationship. Still, future research should strive to include this variable, especially when investigating cyberbullying in a relational context. The regressions for passive observing and active defending explained a relatively low amount of variance, and there are other variables, such as number of bystanders and perceived anonymity, that fit within the BIM but could also be integrated into future studies to better understand how expectancies interact with salient contextual factors to predict bystander behavior.

In addition, participants were asked to recall any incident that occurred in the previous year, which likely meant that they reported more memorable or perhaps severe incidents. Although this factor was somewhat controlled for in that the valence variable tapped into incident severity, it is possible that participants would have been even less likely to intervene (or even remember) less severe incidents.

An adult sample was used, which is a strength in that cyberbullying and online harassment are more frequently examined in middle school and high school students. However, the use of a younger sample may have led to different results in the present study, in that younger individuals may be more acclimated to antisocial online behaviors and thus less likely to see such incidents as an expectancy violation. Future studies should investigate these variables in a younger sample.

Furthermore, the present study was cross-sectional, and results were correlational. Although such research is necessary and ensures ecological validity, future studies could examine EVT in this context using an experimental design. Participants could be exposed to cyberbullying scenarios in which expectancy violation and valence of violation are experimentally manipulated. Such results could triangulate the present findings.

Conclusion

Most adults have observed an incident of cyberbullying or online harassment (Duggan, 2017), but many bystanders do not intervene. The present study extended EVT, the BIM, and research into norms to better understand why a bystander might refrain from taking action when witnessing cyberbullying. The results revealed that participants generally viewed cyberbullying as both an expectancy violation and a negative behavior. When incidents were perceived as negative, bystanders were generally less likely to passively stand by or join in with the perpetrator and more likely to defend the victim. Importantly, those two factors—incident expectedness and incident severity (i.e., negative valence)—interacted to predict passive observing and joining in with the perpetrator. The extent to which an incident violates bystanders’ expectations had a stronger effect on their behavior when the severity of the incident was ambiguous.

References

https://doi.org/10.1037/0022-3514.84.1.18


