The Perpetuation of Online Hate: A Criminological Analysis of Factors Associated with Participating in an Online Attack

James Hawdon
Virginia Tech University
Matthew Costello
Clemson University
Rebecca Barrett-Fox
Arkansas State University
Colin Bernatzky
University of California, Irvine

ABSTRACT

Online extremism, or the use of information technology to profess attitudes devaluing others based on a characteristic such as race, religion, gender, or sexuality, is a growing problem. This has led to myriad harmful effects for some who are exposed to online hate. A critical first step toward stemming the tide of online hate is understanding factors associated with its creation and spread. To that end, this analysis examines factors associated with joining an ongoing attack against a targeted group online. We use insights from four leading criminological theories – routine activity theory, social control theory, general strain theory, and social learning/differential association theory - to investigate who is likely to join an attack on a targeted group when they view such an attack occurring. Using data from a national sample of 15 – 36-year-old Internet users, we conduct an ordinal logistic regression analysis. Results show support for social control theory and strain theory, as low levels of self-control and online strain are both positively correlated with joining an online attack. Similarly, we find support for the applicability of social learning theory; close engagement with online friends and groups is related to an increased likelihood of joining in online hate. Routine activity theory, however, is less relevant for understanding our outcome. Taken together, our findings shed light on factors associated with the perpetuation of online hate, and, in doing so, offer avenues for reducing its growth.

Keywords: Online Hate, Online Extremism, Criminological Theory
Scroll down to the content of the page.
online attack against a group. We then use logistic regression to predict who joins in an attack when they see it online. We conclude by discussing the theoretical and practical implications of our research.

**LITERATURE REVIEW**

A number of theories originally developed to explain involvement in criminal activities in the “real world” have been applied to the virtual world. In general, most tests to determine if our theories are truly general and apply to both the analog and virtual world find that they do indeed predict online criminality and deviance. While some of the theories do not apply directly (see, for example, Yar’s (2005) discussion of RAT), the theories are typically found to apply once modified. However, most of this work investigates behaviors such as online theft and fraud, identity theft, piracy, and sexual harassment. Do our leading theories also apply to participating in hateful online attacks against a group? We will consider RAT, General Strain Theory, Self-Control Theory, and Differential Association Theory.

**Routine Activity and Lifestyle Theories**

Routine activity theory (Cohen & Felson, 1979), which is the most influential theory of victimization (Miró, 2014), argues there are three necessary factors for crime to occur: (a) the presence of motivated offenders, (b) the presence of a suitable target, and (c) the absence of a capable guardian. Crime will be most likely to occur when all three components spatio-temporally converge. Routine activity theory proposes that victimization stems from the “recurrent and prevalent activities” in which individuals are involved, which in turn influence the likelihood that the three necessary factors for crime are present (Cohen & Felson, 1979). Therefore, an individual’s routines influence his or her risk of being victimized.

This perspective has recently been applied to the online world by recognizing that offenders and victims intersect within a virtual network instead of physical space, and virtual contact can occur asynchronously (Leukfeldt & Yar, 2016; Vakhitova & Reynald, 2015;). Thus, online routine activities can increase the likelihood of victimization by bringing potential targets into virtual contact with potential offenders in environments lacking guardians (see Eck & Clarke, 2003; Reens, Henson, & Fisher, 2011). Adapting the theory to the online world, several recent studies successfully explain a variety of types of cybervictimization, ranging from fraud and identity theft to harassment and other forms of cyberviolence (e.g., Bossler & Holt, 2009; Bossler, Holt, & May, 2012; Costello, Hawdon, & Ratliff, 2017; Costello, Hawdon, Ratliff, & Grantham, 2016; Hawdon et al., 2015;
These studies find that engaging in risky online behaviors such as downloading games and music from unknown websites, using file-sharing programs, instant messaging, opening unknown email attachments, and clicking on pop-up messages increases cyberharassment (Hinduja & Patchin, 2009; Holt & Bossler, 2008; Marcum, 2009; Marcum, Higgins, & Ricketts, 2010; Navarro & Jasinski, 2012; Navarro & Jasinski, 2013; Pratt, Holtfreter, & Reisig, 2010; Reyns, 2013; Reyns & Henson, 2015; Reyns et al., 2011; van Wilsem, 2011). Similarly, those who anonymously confide in others online experience greater levels of cybervictimization than those who are more guarded in their online behaviors (Hawdon et al., 2014; Reyns & Henson, 2015). Holt and Bossler (2008) found that general computer use, such as playing video games, spending time in chatrooms, online shopping, or checking email increased the likelihood of experiencing cyberviolence. One of the most robust findings regarding online routines and victimization is that the use of social networking sites increases the likelihood of victimization (Bossler & Holt, 2009; Bossler, Holt, & May, 2012; Costello et al., 2016; Hawdon, et al., 2014; Leukfeldt & Yar, 2016; Navarro & Jasinski, 2012; Reyns et al., 2011; van Wilsem, 2011).

In addition to behaviors that can bring one into virtual contact with online offenders, other factors can increase victimization by reducing guardianship or collective efficacy. Guardianship is “the presence of a human element which acts—whether intentionally or not—to deter the would-be offender from committing a crime against an available target” (Hollis, Felson, & Welsh, 2013, p. 76). Collective efficacy includes working trust and social ties within communities as well as the willingness to intervene to achieve social control. The findings with respect to guardianship and cybercrime are inconsistent (e.g., Bossler & Holt, 2009; Leukfeldt & Yar, 2016; Reyns, 2015), in part due to conceptual uncertainty across both studies and types of victimization (Vakhitova & Reynald, 2015). Recently, Hollis, Felson, and Welsh (2013) explicitly rejected the notion that guardianship is social control; instead, guardianship occurs when the mere presence of a person or persons acts to reduce the likelihood of a crime occurring (also see Felson 1998). Finally, in addition to online social control and guardianship, the use of target-hardening devices such as antivirus programs, firewalls, filtering and blocking software can potentially reduce cybervictimization. In general, researchers have found that target hardening has very little effect on violent cybervictimization (e.g., Holt & Bossler, 2008; Marcum, 2009; Marcum, Higgins, & Ricketts, 2010).

Although RAT is a theory of victimization, the well-documented overlap between being victimized and engaging in victimizing behaviors (see
Jennings, Piquero, & Reingle, 2012) suggests that exposure to online hate materials and joining in the attack of a group would be positively related. Indeed, this relationship could be partially due to the strain that victimization causes, and this potential linkage is best explained through Agnew’s (1992) general strain theory (GST).

**General Strain Theory**

GST posits that stressful life events produce negative effects (e.g., anger, frustration, or sadness) that can lead to delinquent coping responses. Agnew (1992) identifies three types of strain: failure to achieve positively valued goals; removal of positively valued stimuli; and, presentation of negatively valued stimuli. Strain is linked indirectly to aggression and engagement in violence as well as other problems behaviors because strain also produces negative affect, including feelings of anger, frustration, or sadness (Agnew, 1992). For example, victimization is a negative stimulus, and cyberviolence victimization would be a form of strain that could produce negative emotional states that result in participation in cyberviolence. As noted above, there is a well-documented relationship between victimizing others and being victimized, and this relationship holds online as well as offline (see Bossler & Holt, 2009; Costello et al., 2016; Holt & Bossler, 2008; Holt & Bossler 2013; Jennings, Piquero, & Reingle, 2012; Marcum, Higgins, & Rickets, 2014; Ngo & Paternoster, 2011; Reyns et al., 2011). For instance, participating in sexting activities increases with threats of violence and other forms of cybervictimization (Reyns, 2013), and those who see expressions of cyberviolence and extremism are more likely to produce online hate materials than are those who are not exposed to these materials (Costello & Hawdon, 2018; Hawdon et al., 2014). Recent work by Costello and Hawdon (2018) shows that online users who have been targeted by online hate are more than eight times as likely to produce such material, relative to those who have not been directly targeted. Further, use of general message boards and Reddit, a news aggregation and discussion site where hate is pervasive, was found to correlate with the production of online hate. This reciprocal relationship between engaging in cyberviolence and being a victim of cyberviolence may result from victimization creating strain.

Indeed, research demonstrates that cyberviolence victimization can be a strain-inducing experience and result in committing acts of cyberviolence both directly and indirectly (Ak, Özdemir & Kuzucu, 2015; Bae, 2017; Hay, Meldrum, & Mann, 2010). Cyber-aggression by one (or a group) upon another intuitively involves the presentation of negatively valued stimuli. Moreover, from a GST perspective, reported social acceptance of cyberviolence by others is problematic because people seek affirmation and
approval from their peers. Cyberviolence, however, derails the pursuit of social affirmation and approval because of the complexities with rejection and exclusion associated with cyber-aggression. Researchers argue that when individuals perceive themselves to be rejected or otherwise socially excluded, a number of emotional, psychological, and behavioral ill effects can result (Keipi et al., 2017). In other words, the failure to achieve peer acceptance as signaled through cyber-aggression and victimization may produce stressful feelings that ultimately result in participating in cyberviolence.

**General Theory of Crime/Self-Control Theory**

Gottfredson and Hirschi’s (1990) general theory of crime or self-control theory may also explain involvement in cyberviolence. Simply put, Gottfredson and Hirschi (1990) argue that criminals and deviants lack the ability to regulate their behavior. That is, they lack self-control. Claiming that levels of self-control are determined early in life and remain invariant over the life-course, Gottfredson and Hirschi (1990) argue that the correlations among deviant behavior, dangerous-but-legal behaviors such as smoking and crime are so high because these are all manifestations of the same lack of self-control. Similarly, the well-documented correlations between crime and a host of individual traits and characteristics such as intelligence, educational attainment, divorce, drug use, and a host of other problems are due to these being “manifestations of low self-control.” In short, “people who lack self-control will tend to be impulsive, insensitive, physical (as opposed to mental), risk-taking, short-sighted, and nonverbal, and they will tend therefore to engage in criminal and analogous acts” (Gottfredson & Hirschi, 1990, p. 90). At its core, self-control theory is a variant of rational choice theory because those with low self-control are unlikely to calculate properly the negative outcomes of their behavior. As Gottfredson and Hirschi (1990) say,

So, the dimensions of self-control are, in our view, factors affecting calculation of the consequences of one’s acts. The impulsive or short-sighted person fails to consider the negative or painful consequences of his acts; the insensitive person has few negative consequences to consider; the less intelligent person also has fewer negative consequences to consider. (p. 95)

Thus, those with low self-control would likely emphasize the immediate rewards associated with cyberviolence and fail to recognize the potential dangers associated with the behavior. Indeed, for some circulating online hate has been found to be liberating; the Internet offers an outlet for purvey-
ors of hate to spread their socially-undesirable views and vent frustrations (Douglas, 2007). According to psychologist Bernard Golden (2016), participating in group hate can fill a void in individuals lacking a sense of identity. The dissemination of hate can distract one from feelings of powerlessness or inadequacy, while simultaneously fostering connection with likeminded individuals.

The general theory of crime has produced numerous attempts to test it and various assertions made by Gottfredson and Hirschi (1990). In general, these tests have been favorable although several authors (e.g., Geis, 2000; Higgins, 2006; Holt, Bossler, & May, 2012; Pratt & Cullen, 2000; Wikström & Svensson, 2010) note or find that some of the more general claims of the theory are somewhat limited. Nevertheless, there is empirical support for the claim that low self-control is related to participation in a variety of cybercrimes (e.g., Bae, 2017; Clevenger, Navarro, & Jasinski, 2016; Donner, Li, Holt, Bosler, & May, 2016; Marcum, & Jennings, 2014; Marcum, Higgins, & Rickets, 2014).

**Social Learning / Differential Association Theory**

Social learning theories, which evolved from differential association theory (Sutherland & Cressey, 1974), suggest subgroup variation in attitudes toward violence. Accordingly, individuals learn antisocial values and techniques through intimate social relations, especially family and friends. Because some individuals and groups have positive attitudes toward violence or justify violence under particular circumstances, social learning theorists assume youth may be bonded to others while simultaneously holding attitudes favorable to law violation. While the original conceptions of social learning theory postulate that violence stems from individuals learning pro-violence definitions and attitudes within interpersonal relationships, more recent conceptualizations also include definitions of learning from behavioral psychology. Mechanisms such as imitation and personal and vicarious reinforcement are also powerful means for learning violence (Akers, 1977). Thus, violence results from continual and reciprocal processes of social observation, attitude internalization, and real and perceived reinforcements from self and others. That is, individuals learn to be violent through interactions with others who define their violent behaviors positively, reward that behavior, and help them internalize pro-violent orientations.

Social learning/differential association theory is one of the most widely tested and supported theoretical perspectives of crime (see Pratt et al., 2010), and a growing body of literature demonstrates its applicability to online settings. For example, among a sample of university students, Hollinger (1992) found that friends’ involvement in computer piracy signifi-
cantly increased respondent involvement in piracy. Similarly, Skinner and Fream (1997) report that associating with friends who participate in computer crime is the strongest predictor of engaging in piracy, accessing or trying to access a computer account, changing another’s computer files, or creating or using a virus. Testing several criminological theories, including strain theory, techniques of neutralization, social learning theory, and self-control theory, Morris and Higgins (2009) found that differential association was the most pronounced theoretical predictor in self-reported piracy. Finally, Aker’s (1977) social learning theory—an elaborated version of differential association theory—has been supported in a number of studies of cybercrime (Higgins & Makin, 2004a; Higgins & Makin, 2004b; Higgins et al., 2006; Ingram & Hinduja, 2008). While social learning/differential association theory has been applied to several forms of cybercrime, research on its applicability for cyberviolence has been somewhat limited. Yet, Hinduja and Patchin (2013) found youth who were punished by their parents or adults at school for engaging in cyberbullying were less likely to engage in cyberbullying. Similarly, researchers have found that social learning variables predict sexting and cyberbullying (Li et al., 2016; Marcum, Higgins, & Rickets, 2014). While we are unaware of any empirical work that uses social learning theory to predict the creation or dissemination of online hate, Hawdon (2012) has theoretically outlined how social learning theory applies to creating and disseminating extremist materials.

DATA AND METHODOLOGY

This analysis assesses factors associated with seeing online hate material and joining in the hate. We focus on the potential explanatory power of four major criminological theories – routine activity theory, general strain theory, self-control theory, and social learning/differential association theory. The analysis begins by presenting descriptive attributes of the data. This is followed by an ordinal logistic regression analysis. This technique is used because our outcome of interest is ordered and categorical. The effect of independent variables are reported as odds ratios, which show relative changes in the odds of an outcome when an independent variable’s value is increased by one unit, holding all other effects constant.

Sample

We use data from a sample of 900 American Internet users between the ages of 15 to 36. The data were collected during the week of November 21, 2016 from demographically balanced panels of people who agreed to participate in surveys. Survey Sample International (SSI) administered the
panels. SSI recruit panel members through various permission-based techniques, including random digit dialling and banner ads, and they provide incentives to panel members for participating in and completing surveys. To recruit respondents, SSI emailed invitations to a sample of panel members between the ages of 15 and 36. In addition to age, the sample was stratified to reflect the U.S. population on gender and geographic region. The ages 15 to 36 were selected because these data are from a study of online hate designed to provide comparative samples from earlier research conducted in several European nations (e.g. Räsänen et al., 2016).

Demographically balanced panels protect against bias in online surveys. Screening can eliminate respondents and panelists who have previously participated (Evans & Mathur, 2005; Wansink, 2001), and the recruitment and selection processes, the use of pre-panel interviews and incentives increase the respondents’ seriousness and attention completing the survey thereby improving the validity of responses (see Wansink, 2001). Similar samples have been used in several studies related to cyberhate (e.g. Costello, Hawdon, & Ratliff, 2017; Costello et al., 2016; Näsi, Räsänen, Hawdon, Holkeri, & Oksanen, 2015; Näsi et al., 2014; Räsänen et al., 2016).

**Dependent Variable**

The dependent variable asks respondents to indicate how often they join in mean or offensive behavior on social networking sites when they encounter it. Respondents decided what constituted mean or offensive behaviors. They “joined in” the behavior if they engaged in similar behavior or explicitly approved of the materials they encountered online. Potential responses range from “never” to “frequently.” A majority of the respondents (58.7%) indicate that they “never” engage in such behavior. Smaller shares responded that they do so “only once in a while,” (18.1%), “sometimes,” (17.4%), or “frequently” (5.8%).

**Independent Variables**

RAT emphasizes the intersection of proximity to motivated offenders, suitability of targets, and a lack of capable guardianship to explain crime. While RAT is typically used to explain crime in the physical world, we use a modified version that allows for application to an online setting. Recent scholarship has demonstrated the applicability of RAT to understand online crime and deviance.

Proximity, or exposure, to online offenders is assessed using measures of social networking site (SNS) usage, and hours per day online. We expect
that increased proximity to online hate will afford individuals more opportunities to engage in hateful behavior and hence be positively associated with our outcome variable. Countless SNS exist, and some have reputations for being particularly hateful. Notably, recent work found that people are more likely to be exposed to online hate on YouTube, photo-sharing sites like Snapchat, and Tumblr (Costello et al., 2016). Further, users of general message boards and Redditt were found to be more likely to produce hateful material online (Costello & Hawdon, 2018). Thus, there is reason to believe that individual SNS could inspire various types of hate using diverse forms of presentation. We therefore initially investigated the potential individual effects of SNS separately. None of these SNS achieved statistical significance after relevant control variables were included in the regression model, however, and we therefore opted to measure SNS usage dichotomously. While the particular social media platform that individuals use may not matter per se in this analysis, including an overall measure of SNS usage is theoretically important as an online routine, since prior work shows that it predicts exposure to online hat material (Costello et al., 2016; Oksanen et al., 2014). Our dichotomous measure captures individuals who use a high volume of SNS. Respondents who report using 15 or more SNS are scored as a “1”, while those who use 14 or less are scored as a “0.” Only 7.9% of respondents use more than 15 SNS.

We asked respondents how many hours per day they spend online to assess time online. The variable response set ranges from 1, “less than one hour per day,” to 6, “ten or more hours per day.” Respondents spent between three and five hours online per day, on average. We assess guardianship by looking at the living arrangements of respondents. Living arrangements is measured dichotomously, with respondents indicating if they live alone or with friends or family members. Only 9.6% of our sample lives alone. We expect respondents with less guardianship, or those who live alone, to be at a heightened risk of encountering, and perhaps engaging in, online hate. Since this work focuses on participating in hate, not victimization, we do not examine indicators of target suitability, which are relevant for studies of victimization.

Social learning theory suggests that offenders learn their criminal or deviant behaviors from intimate contacts who then reward and reinforce that behavior. It is therefore likely that factors that account for exposure to online hate may similarly account for its perpetration. This might be especially true regarding factors that increase proximity to motivated offenders online. These indicators of exposure can in turn indirectly lead to the perpetration of hate by bringing individuals into contact with offenders who recruit, groom, or reward those who join in their hateful behavior. We control for several variables that represent characteristics of online users who
would likely comprise the group of offenders apt to encourage others to join in their deviance, and therefore represent agents of social learning.

First, we use a measure that asks respondents how frequently they see hate material online. This measure was assessed using a 4-point scale, ranging from 1, corresponding to “never” to 4, corresponding to “frequently.” Most respondents reported seeing hate “sometimes” (40%) or “only once in a while” (29.3%). Smaller shares indicated seeing hate material “frequently” (21.4%) or “never” (9.6%). Second we control for closeness to an online community. Closeness to an online community is measured on a 5-point scale that ranges from 1, or “not at all close,” to 5, or “very close.” Most respondents report a moderate to high sense of closeness to an online community, with 28.1% answering with a “4” and 27.6% responding with a 3. It was least common for respondents to say that they are “not at all close” (10%) to an online community. Third, we control for interactions with close friends online, which is measured with a 4-point scale with responses ranging from 1, or “almost none of them,” to 4, or “almost all of my close friends are online.” Half of our sample (49.4%) said they interact with a “few” online friends, but that “most of their close friends they also see offline.” It was less common for respondents to say that they interact with “almost none” (14.6%) or “almost all” (8.6%) of their close friends online. We expect respondents who see online hate more frequently, have closer ties to an online community, and interact with more close friends online to be more likely to join in hateful behavior online because hateful ideologies are likely to be nurtured and reinforced in tightly bonded groups who hold hateful beliefs (see Hawdon 2012). We lack direct measures of being involved in groups that advocate hate, however, so these measures serve only as indirect proxies of involvement with deviant peers.

We also control for demographic traits of respondents that might bring them into contact with hateful offenders. Notably, the realm of online hate is currently dominated by rightwing hate that champions white-supremacy, patriarchy, and nationalism (Hawdon et al., 2014; Potok, 2015; Ratliff et al., 2015). We therefore control for gender, race/ethnic minority status, immigrant status, and political ideology. Gender and minority status are measured dichotomously. Fifty-three percent of our sample is comprised of men, and 82% is white. Immigrant status is assessed by asking respondents if their parents were born outside of the United States. Nearly one third (31.1%) indicated that their parents were. Political ideology is assessed using a 7-point scale that ranges from 1, or “extremely liberal,” to 7, “extremely conservative.” The largest share of our sample (29.6%) report moderate political views. One-third (33%) identify as extremely liberal or liberal, and 20.3% categorize their political ideology as extremely conservative or conservative. We expect white men/boy respondents whose parents
were born in the United States, and conservatives to be more likely to join in online hate, given the dominance of far-right extremism online. However, it is important to note that not all respondents who fall into these categories are more likely to engage in such behavior. Indeed, the likelihood is probabilistic, and based on the notion that they are merely more likely to hold worldviews that align with rightwing hate.

We evaluate self-control theory using a composite of three indicators that approximate self-control. Examining the factorability of the measures demonstrates that each item shares variance with the other two items and the combined index explains 66.8% of the total variance in the items. The first measure asked respondents to assess on a 1-to-10 scale how true the statement “I enjoy taking risks” was for them. Higher scores indicate a higher proclivity for risk-taking. The average level of risk taking in our sample is 6.8. Nearly half of the respondents (46.18%) answered with a score of 8 – 10. The second measure asked respondents to rate the accuracy of the statement “I often do things that feel good in the moment, but I regret later on.” Possible responses ranged from 1 to 10, with higher scores denoting that the statement was truer of them. Respondents reported an average score of 5.5, and nearly identical shares of our sample responded with a score of 8 – 10 (29.89%) and 1 – 3 (29.55%). The final component of this measure, using the same 10-point scale, asked survey-takers to rate the statement “sometimes I can’t stop myself from doing things my friends are doing, even if I know it is wrong.” The average response for the survey was 4.6, and a large share of respondents, 43.84%, indicated that this is not very true of them, responding with a score below 3. We expect individuals who show less self-control to be more inclined to join in hateful behavior online, in line with extant work showing that low self-control correlates with the engagement in various types of cybercrime.

We examine strain using two measures. The first looks at online strain, which represents the presentation of a negatively valued stimulus. This variable is an additive indicator composed of two dichotomous measures: the first asks respondents if they have ever been the target of online hate, and the second asks if they have ever been victimized by an online crime. These two measures are highly correlated ($r = .52$). Over 30% of the sample said they were a victim of online hate, and 18.6% of respondents reported they had personally been the victim of an online crime. The second measure evaluates offline strain, looking at economic engagement. This variable represents failure to achieve a positively valued goal. We use an indicator that categorizes individuals who are in school or working full-time as economically engaged, and those who are unemployed or only working part-time as not economically engaged. A sizable share of our sample (75.2%) is economically engaged by this definition. We expect individuals who experi-
ence strain to be more apt to join in hateful behavior online, given the established link between strain and delinquency more generally.

Finally, we include indicators of age and education as demographic control variables. Education is measured on a 5-point ordinal scale, ranging from 1, or “less than a high school diploma,” to 5, corresponding to “a master’s degree, professional degree, or higher.” The most common response in our survey was that individuals had a college degree (29.46%). The age range spans 15 – 36 year olds, and the average age of respondents is 24.7 years old. We do not advance specific hypotheses about these three variables.

**FINDINGS**

Table 1 reports the means, standard deviations, and minimum and maximum values for all variables in the analysis. A visual inspection of correlations between the independent variables does not appear to raise concerns over multicollinearity. In fact, the only correlation above .6 is between age and education (.68). A variance inflation factor (VIF) test affirms that multicollinearity is not a concern. The mean VIF score for the full model is only 1.27. A correlation matrix is available upon request.

Table 2 shows the results of regressing the dependent variable, joining in online hate, on the independent variables. We utilize a two-model sequence for this analysis. The first model includes measures that test RAT and social learning/differential association theory. It also includes sociodemographic control variables. The second model adds a measure to examine social control theory and two measures that look at general strain theory.
Model 1 does not support the applicability of our RAT variables to our outcome variable. SNS usage is not significantly related to joining in online hate and, contrary to expectations, time per day online is negatively associated with our dependent measure (OR=.91, $p < .05$). Our measure of guardianship also fails to demonstrate a significant relationship with joining in online hate.

We do, however, find support for our hypotheses regarding differential association. Notably, seeing online hate frequently (OR=1.35, $p < .001$), being close to an online community (OR=1.83, $p < .001$), and interacting with close friends online (OR=1.22, $p < .05$) all demonstrate a positive association with seeing hate and joining in, suggesting potential social learning processes. Men are also more likely than women to engage in such behavior online (OR=1.79, $p < .001$). Contrary to our expectations, though, respondents with foreign-born parents were more likely to see hate and join in (OR=1.71, $p < .001$), and race/ethnic minority status and political ideology were both unrelated to our dependent measure. Additional findings show that level of education is positively associated with seeing hate and joining in (OR=1.31, $p < .001$), though age is not a significant correlate.

Model 2 shows that those with less self-control are more than twice as likely to join in being hateful upon encountering online hate (OR=2.10, $p < .001$). This is the strongest effect in the model, lending robust support to the applicability of social control theory to an online setting. General strain theory receives varied support in this model. Indeed, experiencing online

### Table 1: Descriptive Statistics of All Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean/%</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>See Hate Online &amp; Join In</td>
<td>1.70</td>
<td>0.95</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>High SNS Usage = 1</td>
<td>7.9%</td>
<td>0.27</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hours/Day Online</td>
<td>3.75</td>
<td>1.40</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Live Alone = 1</td>
<td>9.6%</td>
<td>0.29</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Frequency/See Offensive Online</td>
<td>2.73</td>
<td>0.90</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Close to Online Community</td>
<td>3.33</td>
<td>1.23</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Interact with Friends Online</td>
<td>2.30</td>
<td>0.82</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Male = 1</td>
<td>53%</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>White = 1</td>
<td>82.4%</td>
<td>0.38</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Foreign-Born Parent(s) = 1</td>
<td>31.1%</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Political Ideology</td>
<td>3.66</td>
<td>1.76</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Age</td>
<td>24.70</td>
<td>6.80</td>
<td>15</td>
<td>36</td>
</tr>
<tr>
<td>Education</td>
<td>3.15</td>
<td>1.38</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Self-Control (High to Low)</td>
<td>0</td>
<td>1</td>
<td>-1.97</td>
<td>1.89</td>
</tr>
<tr>
<td>Online Strain</td>
<td>0.49</td>
<td>0.74</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Economically Engaged = 1</td>
<td>75.2%</td>
<td>0.43</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
strain is positively associated with joining in online hate (OR=1.73, \( p < .001 \)), though economic engagement is not significantly related to our dependent measure. The results from the first model remain mostly intact, with the exception of the positive effect of having parents born outside of the U.S., which fails to reach significance in this model.

**TABLE 2: Ordinal Logistic Regression Analysis of Encountering Online Hate Material and Joining in the Hate**

<table>
<thead>
<tr>
<th>See Hate Online &amp; Join In</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio</td>
<td>Std. Error</td>
</tr>
<tr>
<td>High SNS Usage = 1</td>
<td>1.17</td>
<td>0.29</td>
</tr>
<tr>
<td>Hours/Day Online</td>
<td>0.91*</td>
<td>0.05</td>
</tr>
<tr>
<td>Live Alone = 1</td>
<td>0.86</td>
<td>0.24</td>
</tr>
<tr>
<td>Frequency/See Offensive Online</td>
<td>1.35***</td>
<td>0.12</td>
</tr>
<tr>
<td>Close to Online Community</td>
<td>1.83***</td>
<td>0.14</td>
</tr>
<tr>
<td>Interact with Friends Online</td>
<td>1.22*</td>
<td>0.12</td>
</tr>
<tr>
<td>Male = 1</td>
<td>1.79***</td>
<td>0.29</td>
</tr>
<tr>
<td>White = 1</td>
<td>0.97</td>
<td>0.22</td>
</tr>
<tr>
<td>Foreign-Born Parent(s) = 1</td>
<td>1.71***</td>
<td>0.28</td>
</tr>
<tr>
<td>Political Ideology</td>
<td>1.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Age</td>
<td>0.98</td>
<td>0.01</td>
</tr>
<tr>
<td>Education</td>
<td>1.31***</td>
<td>0.10</td>
</tr>
<tr>
<td>Self-Control (High to Low)</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Online Strain</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>Economically Engaged = 1</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td><strong>LR X2</strong></td>
<td>185.49</td>
<td></td>
</tr>
<tr>
<td><strong>Log Pseudolikelihood</strong></td>
<td>-763.71</td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>763</td>
<td></td>
</tr>
</tbody>
</table>
* \( p < .05 \); ** \( p < .01 \); *** \( p < .001 \) (two-tailed tests)

**DISCUSSION**

While most people who witness cyberhate do not join in, a notable minority, 23.8%, do at least sometimes, suggesting that the threshold for participating in online hate is something that many people are willing to cross under certain circumstances. When a person engages in such behavior, doing so increases risks for them, in addition to the harm it causes the targets of their hate. Because online exposure to - and engagement in - hate is related to offline hateful behavior (Cowan & Mettrick, 2002; Foxman & Wolf, 2013), addressing digital hate may prevent harm from being done both online and in the real world.
Routine activity theory proposes that increased exposure to motivated offenders and lack of capable guardianship should correlate with seeing hate and joining in. However, neither more time spent online each day nor increased SNS usage increases the likelihood of our sample participants joining in an online attack. Living alone is likewise unrelated to joining in hate. Even so, while the number of SNS visited and time online are not positively related to a person’s decision to join an online attack, the kinds of places visited and people with whom respondents observe or engage are, lending support to the power of the modeling of social behavior.

Indeed, our findings suggest that those who see hate frequently online, are close to an online community, and interact with close friends online are more likely to join in hateful behavior. All of these findings lend credibility to the application of social learning/differential association theory to studying online deviance. Being online with others who one feels close to may prompt deviant behavior if those online acquaintances engage in or promote deviance. Unfortunately, parental disapproval of online friend’s parents labeled “deviant” has been shown to increase children and teens’ involvement with those very people online (Keijzers et al., 2011), producing the exact opposite effect that parents and other guardians desire. In short, young people seek out and engage in online socialization with peers who engage in deviance.

We also find that men are more likely to join in online hate. This is unsurprising for a few reasons. First, this finding aligns with studies of offline deviance in general (Tedor, 2015), offline hate in particular (Ferber, 2004), and other kinds of online deviance (Abougaoude, Savage, Starcevic, & Salame, 2015; Donner, 2016) that show men partaking in deviance more readily than women. Second, in this particular case, the positive male effect might also be a byproduct of the type of hate that presently permeates cyberspace. Rightwing hate has a virulent misogynic and anti-feminist streak. Thus, if, statically individuals are most likely to see rightwing hate online, it would stand to reason that men, not women, would most likely participate in the hateful behavior.

Further investigation into the nature of online relationships—if they are with people who advocate or warn against online deviance—can illuminate whether and how online socialization teaches cyberhate. Studies of these online relationships and how people enter them can help scholars understand whether people prone to hateful online behaviors flock together (and perhaps to particular SNSs that may be more accommodating of their views and behaviors) and thus reinforce hateful behaviors or if, alternatively, they find each other prior to being cyberhaters and then develop hateful views that in turn inspire hateful behaviors, a framework termed feathering (Costello et al., 2016; Hirschi, 2017; Sutherland & Cressey,
of cyberhate socializations, and further research could distinguish when each model is most applicable.

The strongest finding of this analysis is that low levels of self-control predict a higher likelihood of engaging in deviant behavior. This may be because low levels of self-control are an indicator of a poor ability to foresee unpleasant consequences for oneself or others, an insensitivity to unpleasant consequences, or the ability to make accurate predictions and/or feel their consequences but not to regulate behavior accordingly. In an online setting, this could take the form of a cyberhate attackers not understanding the pain they cause others, not understanding the potential negative social consequences of their comments (including, for example, loss of job or public shaming), simply not caring about those consequences, or accurately judging and caring but exercising poor impulse control. Additionally, people with poor impulse control could flock together, socializing members into a further lack of consideration of consequences, impulsivity, or cruelty. Finally, the culture of particular SNSs may encourage such impulsive behavior through, for example, upvotes that allow readers to reward the most hateful comments, or loose or non-existent moderation, so hateful comments go without a formal sanction from the group.

Finally, general strain theory received mixed support. Economic uncertainty is commonly cited as a primary source of strain that can lead to deviant or criminal behavior. Our results do not show evidence such a relationship, though. This could be attributable to the lack of applicability of economic strain to an online setting, or it could indicate the imprecise nature of our measure, which primarily taps into whether a person is employed or unemployed. Given that a large share of our sample is college-aged or younger, economic engagement may not be a primary concern for a large swath of our respondents. Interestingly, online strain, which represents the introduction of a negative stimulus into a person’s life, is associated with engaging in online hate, though. That is, individuals who have been targeted by hate online or fallen victim to an online crime demonstrate an increased likelihood of engaging in online hate. This speaks to oft-found relationship between being victimized and victimizing (e.g. Bossler & Holt, 2009; Bossler, Holt, & May, 2012; Costello, Hawdon, & Ratliff, 2017; Marcum, Higgins, & Ricketts, 2014).

Study Limitations

This study has a few limitations that require discussion. First, our sample is comprised of individual between the ages of 15 and 36. This age range was selected because youth and young adults are avid Internet users,
and therefore might be more likely to regularly encounter online hate. Even so, our results cannot be generalized to older and younger individuals, who are increasingly spending more time online. Second, our study relies on respondents’ perceptions of hate material. While we provide survey-takers with a definition of hate, it is ultimately up to them to interpret what they see online. And, to be sure, not everyone has a parallel definition of hate. Finally, our measures of social learning/differential association are indirect. We lack explicit measures of the type of people and groups that our respondents interact with online. Having such measures would allow for a more precise test of the effects of social learning on joining in online hate.

**CONCLUSION**

Using data from a national sample of 15- to 36-year-old Internet users, this study sought to adjudicate between leading criminological theories—routine activity theory, general strain theory, self-control theory, and social learning/differential association theory—by predicting which factors are associated with joining in attacking a targeted group when such attacks are seen online. Results from our ordinal logistic regression analysis provide strong support for self-control theory, mixed support for social learning theory and general strain theory, and a lack of support for routine activity theory. Most notably, individuals who reported experiencing online strain, less self-control, higher frequency of exposure to offensive content, and closeness to an online community were all more likely to join in online hateful behavior against a targeted group. With these findings in mind, we now outline actionable recommendations that may assist in the identification, mitigation and prevention of the creation and spread of online hate.

Given that a lack of self-control increases the propensity to engage in hateful activity online, we foresee at least two strategies that could curb the proliferation of such behavior. First, social networking sites and other online communities can take steps to implement more robust moderation in order to remove hateful content more quickly. Not only would this approach reduce the impact on the groups targeted by hate by reducing their opportunity to be exposed to the hateful content, but it would also reduce the likelihood of galvanizing other individuals who may be sympathetic to the offensive content and wish to join in. The more time a hateful message remains online, the more likely it is to be seen by users, which may produce strain-inducing feelings and eventually compel those individuals to engage in cyberhate in the future. Prominent social networking services such as Twitter, Facebook, and Reddit can play a major leadership role in this regard, as moderation of such sites remains thoroughly lax despite the fact that most SNSs have clear, articulated policies and community standards
regarding the conduct of its userbase. Moreover, the sheer popularity of these sites makes moderation more difficult—thousands of tweets are posted to Twitter every second, for example—meaning that SNSs typically rely on their users to report hateful material as they see it, a well-meaning policy that has the negative consequence of ensuring that users will be exposed to hate in the first place. While many of these sites have vowed to be more diligent when policing for hate materials, the task ahead of them is truly daunting.

Second, social networking services can improve the accountability and transparency of their userbase by reducing anonymity, thus bringing the online and offline world into closer contact. This tactic may sufficiently deter some individuals from engaging in hateful activity when they encounter it online by raising the costs and consequences of poor self-control. As indicated by literature in social psychology (see Svensson, Pauwels, & Weerman, 2017), the anticipation of social stigma and shame can have a deterring effect on an individual’s likelihood to engage in criminal behavior. The online outing of participants who participated in the August 2017 tiki torch-laden “Unite the Right” rally in Charlottesville, Virginia led to several of the named protestors expressing regret for the hateful optics of their demonstration (Ryland, 2017). By making the consequences of participating in hateful behaviors online more salient to users, individuals may think twice before joining in hateful activities and will reduce the overall prevalence of hate on SNS domains. Of course, limiting anonymity would also reduce what many consider to be the appeal of the Internet and a more open exchange of ideas.

Finally, the findings that closeness to an online community and interacting with close friends online increases the likelihood of participating in hateful behavior online deserves further investigation. Indeed, it should be noted that while these factors did not mediate the effects of self-control in this study, the variables we use to assess social learning/differential association are merely proxy measures. The fact that these proxy measures are significant predictors of joining in online hate attacks suggests that online contexts can exacerbate offensive material if one associates with an online community that sanctions and produces hateful content (see Hawdon, 2012). Consequently, scholars need to better understand the subcultures of the communities in which cyberhateful individuals spend their time, and future research should attempt to operationalize involvement in these groups more directly. As has been documented in offline settings (Eliasoph, 1998), these communities often function as spaces for playful but offensive banter. These collective norms of playful or ironic engagement in online hate can serve as a means for generating social solidarity but also as an end in and of themselves (e.g. a genuine expression of attitudes), and the dis-
tinction between these boundaries is frequently quite blurry. Ultimately, hateful behavior is not born, but made, and it is the prerogative of scholars and SNS stakeholders alike to recognize and explain how factors of exposure, strain and self-control can contribute to hateful behavior; disrupt flocking and feathering processes among hateful communities online; and finally, to offer would-be participants of cyberhate alternative sources of social connectedness and belonging.

REFERENCES


Costello, M., & Hawdon, J. (2018). Who are the online extremists among us? Sociodemographic characteristics, social networking, and online experiences of those who produce online hate materials. *Violence and Gender, 5*(1), 55-60.


