Scott, G. G., Brodie, Z. P., Wilson, M. J., Ivory, L., Hand, C. J. and Sereno, S. C. (2020) Celebrity abuse on Twitter: the impact of tweet valence, volume of abuse, and dark triad personality factors on victim blaming and perceptions of severity. Computers in Human Behavior, 103, pp. 109-119.

http://eprints.gla.ac.uk/205894/

Deposited on: 30 October 2020

**Celebrity abuse on Twitter: The impact of tweet valence, volume of abuse,**
**and dark triad personality factors on victim blaming**
**and perceptions of severity**

**Abstract**

Celebrities are increasingly utilizing social media platforms to establish their brand and interact with their fan base, but in doing so they often become the targets of online abuse. While such abusive acts are known to cause severe consequences in the general population little is known about how celebrity abuse is perceived by observers. This study investigated observers' impressions of the severity of online abuse on Twitter, the blame attributed to celebrities for the abuse they received, and the role of the dark triad of observers' personality factors (Machiavellianism, narcissism, and psychopathy) in these decisions. We manipulated celebrity tweet content (negative, neutral, positive) and the volume of abusive comments (high, low) the tweets received. Celebrities received more blame the more negative their initial tweet was, and incidents were perceived as least severe following a negative tweet with a high volume of abuse. Observer impressions were influenced by their dark triad personality factors. Following negative tweets, as observer narcissism increased, victim blame increased and perceived severity decreased. Following positive tweets, as observer psychopathy increased, perceived severity decreased. Results are discussed in the context of the Warranting Theory of online impression formation and the ramifications for celebrity social media use is explored.

## 1. Introduction

Many celebrities and public figures utilize Twitter to generate publicity and build their brand (Marwick & Boyd, 2011). However, they are increasingly the targets of trolls and victims of online abuse (e.g., Garde-Hansen & Gorton , 2013). Abuse often involves multiple perpetrators and can relate to many different topics (Lumsden & Morgan, 2017). In this study we investigated how abuse towards celebrities on the social media platform Twitter is perceived by observers. Specifically, we measure attributed victim blame (VB) and perceived incident severity after manipulating the content of the original celebrity tweet and the volume of abuse received. We also explore the role played by the dark triad of personality factors (Machiavellianism, narcissism, and psychopathy; Jones & Paulhus, 2013) on observers' VB and perceptions of incident severity. This is an important topic of investigation. Online abuse has serious consequences for victims but perceptions of, and support for, celebrity (vs. non-celebrity) victims is under-investigated. How such activity is viewed in the public domain could shape online norms and influence how victims of all types are perceived.

### 1.1 Celebrities on Social Media

Social media is growing in diversity, popularity, and influence, particularly among younger people (Allcott & Gentzkow, 2017; Hermida, Lewis, & Zamith, 2014; Villlanti et al., 2017). Facebook is the world's largest social networking site (SNS) with over 2.3 billion active users (Facebook Newsroom, 2019). The site is mainly used to share information (such as photos) with others, to maintain friendships, and to organise social activities (Garcia & Sikström, 2014; Tosun, 2012). Other SNSs are popular and serve more niche functions: Twitter allows users to broadcast 'tweets' of 280 (originally 140) characters and has ~336 million active users; Instagram is used exclusively for photo sharing and has ~800 million

active users; and Snapchat allows sharing of non-permanent photos and stories and has ~191 million active users (Statistia, 2018b,c,d).

While SNSs have gained popularity due to their function of facilitating connections between public users (Ellison, Steinfeld, & Lampe, 2007), many now include celebrity users who capitalize on sites' access and popularity to publicise themselves and connect with their fans (Marwick & Boyd, 2011; Lim, 2017). While Facebook has 'public figures' and 'brand' pages which users can 'like' or 'follow' (rather than making 'friends' with), other social media platforms have been more fully embraced by celebrities seeking to establish an online presence.

Snapchat, in particular, has a user demographic focused on celebrity culture, and its success has become largely dependent on celebrity interaction and endorsement. (99firms, 2019). Stocks of the company reportedly fell by over $1bn after criticism by Rihanna, and the threatened departure of Kylie Jenner, leading markets to fear that the cessation of interaction from these previously popular and active celebrities would lead to many non-celebrity users disengaging with the app (Skinner, 2018; Vasquez, 2018). Twitter is utilized by many celebrities (Lee & Lim, 2016; Thomas, 2014) who regularly tweet to develop their personal brand (Page, 2012) and to create a sense of intimacy and affiliation with their fans (Marwick & Boyd, 2011). Via tweets, celebrities can publicise their opinions and likes (Gayle & Lawson, 2013) and raise awareness of themselves and causes they support (Alexander, 2013) and generate public interest (Wu, Hofman, Mason, & Watts, 2011).

As with all individuals in a digital world, celebrities' images increasingly encompass their online persona, with followers viewing Twitter as an authentic source of celebrity information (Van den Bulck, Claessens, & Bels, 2014). Warranting Theory (Walther & Parks, 2002) hypothesizes that, when forming an impression of any individual, observers rely on two categories on online warrants:  social identity claims (statements made by people

about themselves, e.g., "*I'm exceedingly clever*") and behavioral residue (unintentional evidence of a person's personality; on social media this could include statements by others to or about them). As a result of positive self-presentation online, observers will often attribute more weight to behavioral residue and to negative (rather than positive) information when forming impressions (Walther, van der Heide, Hamel, & Shulman, 2009). This may be especially true in the case of celebrities: if public perception is that celebrities utilize social media as a forum for self-promotion, then observers may be suspicious of their motives and genuineness when posting content, thus placing even more weight on others' comments (behavioral residue) to form impressions.

*1.2 Online Abuse*

As the popularity and diversity of social media have increased, so too have cases of online abuse (Dooley, Pyzalski, & Cross, 2009; Selwyn, 2008). Such abuse can take many forms and patterns of abuse are often differentially classified, e.g., cyberbullying, cyberaggression, cyberharassment, and cyberstalking (Maple, Short, Brown, Bryden, & Salter, 2012; Menesini & Nocentini, 2009), with specific categories sometimes difficult to define (Menesini et al., 2012). Although much of this abuse can include private communication between the perpetrator(s) and victim, it also often manifests in the publicly visible online record (e.g., Facebook timeline or Twitter page). According to the Warranting Theory (Walther & Parks, 2002), these publicly visible abusive messages constitute behavioral residue and, as such, would carry weight in the impressions formed of targets and may contribute to a stereotyped impression based on their content (Walther, 1996; 1997). This may be especially true in the case of celebrity victims whom observers may think are being disingenuous with their online communications in order to self-promote and, thus, may deserve any abuse directed towards them.

Victims of online abuse typically report receiving little or no support from friends or authorities (e.g., Crosslin & Golman, 2014; Maple et al., 2012; Shultz, Heilman, & Hart, 2014). This may be because observers attribute some of the blame for abuse incidents to victims (e.g., Russell & Hand, 2017; Scott, Wiencerz, & Hand, 2018; Shultz et al., 2014). Cyberbullying is often perceived as fair or acceptable when it can be explained by the victim's initial behavior (DeSmet et al., 2012). Victim blaming may occur due to observers' beliefs that the world is a just place where people get what they deserve (belief in Just World Theory; Lerner & Simmons, 1966). Alternatively, victim-blaming may occur as an attempt to increase one's own sense of control of our environment, and what happens to us, by attributing abuse to the victim's own disposition (Defensive Attribution Hypothesis; Shaver, 1970). In cases of online abuse, VB increases as more personal information is disclosed by the victim (Weber, Ziegele, & Schnauber, 2013), putting those who are more active on social media at an increased risk. A recent study by Scott et al. (2018) demonstrated that observers' perception of cyberbullying victims (e.g., VB, perceptions of victim attractiveness) was influenced by the volume of abuse directed towards the victims and whether or not the abuse was generated by a single or multiple abusers.

Consequences of online abuse are potentially serious (e.g., physical and mental health impacts on victims: Hinduja & Patchin, 2010), and public sympathy is often lacking from personal and professional networks (Gahagan, Vaterlaus, & Frost, 2016) because incidents are not perceived as severe and victims are often attributed blame (Weber et al., 2013). Negative outcomes reported by lay-public users include psychological effects such as depression and anxiety (Mechanic, Uhlmansiek, Weaver, & Resick, 2000; Short & McMurray, 2009), forced behavioral and lifestyle changes (van Geel, Vedder, & Tanilon, 2014), and suicide (Hinduja & Patchin, 2010). Anecdotal evidence suggests that celebrity

victims of online abuse are likewise impacted, with some victims leaving social media (e.g., Cohen, 2014).

Celebrity victims, who are likely to disclose personal information in an effort to engage with their audience, are unlikely to receive more sympathy than lay-person Twitter users. There have been high-profile cases of online abuse against celebrity targets. Incidents which gain most publicity often include cyberstalking (e.g., Watt & McLean, 2012), or online racist or homophobic attacks (e.g., Carroll, 2012). Abuse against celebrities on Twitter is common; it is often so severe that celebrity targets of abuse choose to deactivate their Twitter accounts (Cohen, 2014). This is perhaps unsurprising, as celebrity tweets often evoke negative feelings (Van den Bulck et al., 2014) and many celebrities' Twitter followers include hostile 'anti-fans' (Gray, 2003). Indeed, many celebrities may provoke gossip, or negative reactions, by being outspoken and using the site to air controversial views (e.g., Muntean & Petersen, 2009). These tweets are identity claims used by observers in forming impressions and may serve to attenuate any impact of abusive messages on the impression formed. Although identity claims carry more weight than behavioral residue when making judgments about certain personality factors (e.g., confidence; Scott & Ravenscroft, 2017), there is no current evidence that this is the case for incidents of online abuse.

Online abuse against both lay (e.g., Hinduja & Patchin, 2010) and celebrity (e.g., Cohen, 2014) Twitter users has shown to have severe negative consequences. Abuse could have professional as well as personal consequences for celebrities, potentially damaging the brand that they intended social media to enhance. It is important how such abusive acts are perceived by observers as this will impact public sympathy and support for victims, which could serve to mitigate some of the negative effects. An additional impact of celebrity abuse occurring in such a public domain is the effects this could subsequently have on other users. Many celebrities have many millions of followers on

social media (Boyd, 2019). If lay-users witness online abuse perpetrated against them, and do not perceive it as serious or severe for the reasons outlined above, this could shape the norms they form of acceptable and unacceptable online behaviour, which would in turn reduce sympathy for victims of all types.

A further aspect of perceived abuse not yet investigated is individual differences between observers. While all viewers will base the impressions they form on the identity claims and behavioral residue of a target's social media profile, how these are interpreted may differ between viewers.

*1.3 Dark Triad Personality Factors*

Although the population in general underestimates the severity of online abuse and its impact on victims, individuals differ in terms of how abusive incidents are interpreted. Specifically, individuals scoring high in the dark triad (DT) of personality traits – Machiavellianism, narcissism and psychopathy (Jones & Paulhus, 2013) – may be likely to underplay the severity of online abuse and to attribute more blame to victims. All three factors have been associated with both workplace bullying (Baughman, Dearing, Giammarco, & Vernon, 2012) and cyberbullying (Goodboy & Martin, 2015). Machiavellianism and narcissism have both been linked to problematic social media use (Kircaburun, Semetrovics, & Tosintas, 2018).

Machiavellianism is reflected by a manipulative and deceptive nature, a lack of concern with conventional morality, and a lack of interpersonal affect (Deluga, 2001). Narcissism, while primarily reflected by high levels of vanity and self-enhancement tendencies not commonly associated with Machiavellianism (Paulhus & Williams, 2002), is similarly characterized by an exploitative interpersonal style, a sense of superiority and entitlement, and selfishness (Millon & Davis, 1996). Finally, psychopathy reflects several

aversive interpersonal (e.g., callousness, remorselessness) and behavioral (e.g., anti-social behavior, impulsivity) characteristics (Douglas, Bore, & Munro, 2012).

Recent research has identified that those high in Machiavellianism and psychopathy are more likely to engage in trolling behaviours (Buckels, Trapnell, & Paulhus, 2014) and cyberaggression (Pabian, De Backer, & Vandebosch, 2015), and are more inclined to use profane and aggressive language online (Sumner, Byers, Boochever, & Park, 2012). Other research argues that psychopathy is a stand-alone, independent predictor of cyberbullying behavior (Gibb & Devereux, 2014; Goodboy & Martin, 2015) and Facebook trolling (Craker & March, 2016). However, van Geel and colleagues found that while Machiavellianism and psychopathy were related to traditional bullying, they were not significant predictors of cyberbullying (van Geel, Goemans, Toprak, & Vedder, 2017), suggesting some inconsistencies in the current literature. Further, while these studies have explored DT personality predictors of online abuse *perpetration*, none have considered their relation to factors relevant to outsider *observation* of abuse.

Recent research has shown that all three dark triad personality factors influence individuals' cognitions, and their perceptions of social situations. Individuals high in DT factors are higher in levels of Schadenfreude (James, Kavanagh, Jonason, Chonody, & Scrutton, 2014), lower in empathy (Doane, Pearson, & Kelly, 2014; Jonason & Krause, 2013; Jonason & Kroll, 2015), generally hold negative perceptions of others, and utilize fewer cues when making assessments' of others' vulnerability (Black, Woodworth, & Porter, 2014). Those high on psychopathy and Machiavellianism are more likely to perceive social situations as competitive, and those high on narcissism to perceive fewer social restrictions (Jonason, Wee, & Norman, 2015). This demonstrates fundamental differences in how individuals high in DT factors perceive both other individuals and social situations, and suggests they may interpret antisocial online behaviours in a way

which would impact their perceptions of victims and the acts perpetrated against them. Thus, the current study investigates how different types of tweets by celebrities as well as observers' DT personality scores influence attributed VB and perceived abuse severity.

*1.4 The Current Study*

In the current study we investigated how the type of tweet written by celebrities (identity claims; negative, neutral, or positive) and the volume of abusive responses by followers (behavioral residue; low or high) affected participants' attribution of VB to the celebrity and participants' perceptions of incident severity. We also examined whether participants' DT personality traits impacted their victim-blaming and severity-perceptions. We predicted:

$H_1$: An initial negative celebrity tweet will result in more victim-blaming.

$H_2$: A higher volume of abuse will result in higher perceived severity.

H3: Attributed VB will be higher among participants scoring higher in DT personality factors

H4: Perceived severity will be lower among participants scoring higher in DT personality factors.

## 2. Method

*2.1 Design and Participants*

The study utilized a 3 (Celebrity Tweet Valence: Negative, Neutral, Positive) × 2 (Abuse Volume: Low, High) within-participants design. Following the presentation of each celebrity tweet, we measured VB and Perceived Severity (PS). After presentation of all celebrity tweets, we measured participants' DT Personality Traits (Machiavellianism, narcissism, and psychopathy). The study was carried out at a UK university and participants

were 184 Twitter users (146 female; $M_{age}$=22.61 years, $SD_{age}$=5.11). The nationalities of participants were self-reported and grouped as follows: 74.5% British; 22.8% European; 1.7% Asian; and 0.5% each as North American and Australian. All were recruited via adverts online and on-campus and took part either voluntarily or for course credit.

*2.2 Materials*

*2.2.1 Stimuli.* Participants were presented with screenshots of six celebrity Twitter pages featuring a tweet by a male celebrity. Each tweet was followed by six comments from unfamiliar sources (non-celebrity Twitter users). Stimuli were manufactured using Adobe Photoshop. Each page contained the following, in order from top to bottom: the celebrity's name and profile picture (taken from Twitter); the tweet itself; the number of 'retweets' and 'favorites' (the numbers for each of these were counterbalanced); and the six comments. The celebrity tweet was Negative, Neutral, or Positive; of the six replies, either two (Low volume) or four (High volume) were abusive, with the rest being neutral. An example stimulus is presented in Appendix A.

*2.2.2 Norming Study.* An initial norming study was conducted with 28 participants (16 females, $M_{age}$=27.75 years, $SD_{age}$=10.64). These participants were recruited online, voluntarily completed the norming study online, and did not participate in the main experiment. This norming study was conducted to select the celebrity identities, initial tweets, and responses used within the main experiment. Participants were presented with a list of 30 male celebrities and asked to rate each on 7-point Likert-type scales of familiarity ("*How familiar are you with this celebrity?*"; *1 = not familiar at all − 7 = extremely familiar*) and feelings ("*What are your feelings towards this celebrity?*"; *1 = negative – 7 = positive*). This established a baseline for impressions towards the celebrities. The six celebrities selected scored high on familiarity ($M_{familiarity}$=5.22 $SD_{familiarity}$=1.51; i.e., participants were likely to have heard of them and recognize their celebrity status), but neither high nor low on feelings

($M_{feelings}$ = 4.51 $SD_{feelings}$ = 1.36; i.e., even though they were familiar, they did not provoke any strong emotive reactions). A one-sample t-test was conducted to demonstrate that these means are significantly different from the mid-point of the scale [$t(5)$=4.378, $p$<0.01]. The six celebrity identities used in the final stimuli were Robert Webb, Matt Baker, Jamie Oliver, Richard Branson, Philip Schofield, and David Guetta.

Participants in the norming study were also presented with a list of 90 tweets and comments – all of which were taken from Twitter – and asked to rate each on 7-point Likert-type scales of valence (*1 = negative – 7 = positive*), arousal (*1 = not arousing – 7 = very arousing*) and politeness (*1 = abusive – 7 = polite*). The mean valence, arousal, and politeness ratings for the Negative. Neutral, and Positive tweets and comments used in the experiment are presented in Table 1. Negative content was low in valence and politeness and high in arousal; neutral content was neither high nor low in valence or arousal, and high in politeness; positive content was high in all three. Examples of tweets used in the final stimuli included: Positive – "*Be disciplined about doin' the little things for your goals – daily. Consistency adds up to success. #ChaseYourGreatness*"; Neutral – "*Weathers getting chilly. I think summer is over*"; Negative – "*Isn't it annoying that the really illiterate & rude people on Twitter are so fucking stupid that they forgot to kill themselves today.*". A complete list of the Positive, Negative, and Neutral tweets is presented in Appendix B, and all Negative and Neutral comments is presented in Appendix C.

Insert Table 1 about here

*2.3 Measures*

Measures of VB and PS were taken from the study by Weber et al. (2013). VB and PS were measured on four- and two-item scales respectively, using 5-point Likert-type scales (Cronbach's α = 0.936 and 0.722, respectively). An example item from the VB measure was:

*"Did the victim provoke the abuse?"* (*1 = strongly disagree – 5 = strongly agree*) and from the PS measure was: *"How severe was the abuse?"* (*1 = not severe at all – 7 = very severe*).

DT personality factors were measured using the SD3 (Jones & Paulhus, 2013). This comprised 27 items, nine each to measure Machiavellianism, narcissism, and psychopathy, with each utilizing a 5-point Likert-type response. Two of the items for Narcissism, and three of the items from Psychopathy, were reverse scored. Example items for each are: Machiavellianism – *"I like to use clever manipulation to get my way"*; Narcissism – *"I insist on getting the respect I deserve"*; Psychopathy – *"It's true that I can be mean to others"*. Measures of Cronbach's alphas for each measure indicated reliability: Machiavellianism $\alpha =$ 0.733; narcissism $\alpha = 0.777$; psychopathy $\alpha = 0.728$.

*2.4 Procedure*

Participants were tested online using SurveyMonkey. Upon answering an advert, participants were sent a link to one of six questionnaires which presented the profiles in one of six pseudo-random orders. After reading task instructions and indicating consent, participants were first asked to fill in a short demographic questionnaire. For each celebrity tweet, they were asked to form an impression of the tweeter and could view each tweet for as long as they wished. Following each tweet, they completed a questionnaire measuring VB and PS. After responding to all tweets, they were asked to complete the DT questionnaire before being presented with debriefing information. The experiment lasted approximately 20 minutes and ethical approval was granted by the host university's Ethics Committee.

3. **Results**

Three sets of analyses were carried out. First, two separate 3 (Tweet Valence: Negative, Neutral, Positive) × 2 (Abuse Volume: Low, High) repeated-measures analyses of variance (ANOVAs) were conducted on participants' VB and PS ratings. Second, Pearson's

correlations were conducted to identify relationships between DT traits, VB and PS. Third, multiple regression analyses were conducted to determine the predictive value of the DT variables for VB and PS. The mean ratings (with standard deviations) across conditions are presented in Table 2.

Insert Table 2 about here

## 3.1 Effects of Tweet Valence and Abuse Volume

For VB, there was a significant and large main effect of Tweet Valence [$F(2,366)=643.41$, $p<0.001$, $\eta_p^2=0.779$]. Bonferroni follow-up contrasts revealed that significantly more VB was attributed to celebrities following a Negative tweet ($M = 3.73$) than following Neutral ($M = 1.65$) or Positive tweets ($M = 1.50$) [both $ps<0.001$]; VB was also significantly higher following Neutral than Positive tweets [$p<0.005$]. There was no significant main effect of Abuse Volume on VB [$F<1$] and no Tweet Valence × Abuse Volume interaction [$F<1$].

For PS, there was a smaller, highly significant main effect of Tweet Valence [$F(2,366)=38.28$, $p<0.001$, $\eta_p^2=0.173$]. Bonferroni follow-up contrasts revealed that abuse was perceived as being significantly less severe following a Negative tweet ($M = 3.23$) than following Neutral ($M = 3.63$) or Positive tweets ($M = 3.68$) or [both $ps<0.001$]; however, there was no significant difference between Neutral and Positive tweets [$p=0.913$]. There was also a significant main effect of Abuse Volume [$F(1,366)=249.99$, $p<0.001$, $\eta_p^2=0.577$] with higher PS in cases with a High ($M = 3.89$) rather than a Low ($M = 3.13$) volume of abusive replies ($p<0.005$). There was no significant Tweet Valence × Abuse Volume interaction [$F(2,366)=1.52$, $p=0.219$].

## 3.2 Personality, Victim Blame, and Perceived Severity

Pearson's correlations (two-tailed) were conducted to identify relationships between DT traits, VB, and PS in each tweet condition and are presented in Table 3. Several significant relationships were identified, all demonstrating small-to-medium strength, based on Cohen's (1988) standards: small, $r = .1$; medium, $r = .3$; large, $r = .5$. Machiavellianism, narcissism, and psychopathy were all positively associated with VB in the negative tweet condition, indicating that all three traits were only associated with higher VB attribution when the original celebrity tweet was negative. In terms of abuse severity, Machiavellianism showed no relation to PS. However, psychopathy and narcissism were negatively associated with PS in the negative tweet condition. Further, psychopathy was also negatively related to PS in the positive tweet condition. Thus, those high in psychopathy perceived abusive tweets as less severe regardless of the original tweet valence, while those high in narcissism only perceived less severity when the original tweet was negative in nature.

Insert Table 3 about here

*3.3 Regressions*

DT variables that were associated with VB and PS at $p<.10$ at the univariate level were taken forward as candidates for multivariable models, as traditional significance limits (e.g., $p<.05$) often fail to establish significance in variables known to be predictive (Bursac, Gauss, Williams, & Hosmer, 2008). Preliminary analyses indicated that the assumptions regarding multicollinearity, independent errors, non-zero variances, normality, homoscedasticity and linearity were not violated.

First, multiple regression was carried out to determine whether the three DT traits predicted VB in the Negative tweet condition and is presented in Table 4. In this model, narcissism was a significant independent predictor of VB in the Negative tweet condition (small to medium effect), but psychopathy and Machiavellianism were not. This suggests that as narcissism increases, VB following negative tweets also increases.

14

Insert Table 4 about here

A second model was developed to determine whether the three DT traits predicted PS of abuse in the Negative tweet condition and is presented in Table 5. Once again, narcissism was a significant independent predictor of PS in the Negative tweet condition (small to medium effect), but psychopathy and Machiavellianism were not. This suggests that as narcissism increases, PS of abuse following negative tweets decreases.

Insert Table 5 about here

A final single linear regression was carried out to clarify the predictive nature of the relationship between psychopathy and PS in the Positive tweet condition and is presented in Table 6. This model demonstrated that psychopathy was a significant predictor of PS in the Positive tweet condition (small effect). This suggests that as psychopathy increases, PS of abuse following Positive tweets decreases.

Insert Table 6 about here

## 4. Discussion

This study investigated the impact of celebrity tweet valence and volume of abusive responses on observers' attribution of victim blame (VB) and perceived severity (PS) of abuse; furthermore, the role of DT personality factors in observers' responses was investigated. We found that the valence of celebrity tweets influenced both attributed VB and PS. Celebrities were blamed most if they had initially tweeted negative content, and least if they had tweeted positive content. Incidents were perceived as least severe following a negative tweet. Volume of abuse only impacted PS, with incidents perceived as more severe when there was a high volume of abuse. When examining the role of participant DT personality factors on their responses to negative tweets, we found that, as narcissism increased, attributions of VB increased, and PS decreased. When abuse followed an initial positive tweet, as participant psychopathy increased, PS decreased.

15

*4.1 The impact of tweet valence and abuse volume*

Our results demonstrate a difference in the impact of identity claims (celebrity tweets) and behavioral residue (volume of abuse) on observers' interpretation of online abuse incidents. Attributed VB was dependent exclusively on identity claims while PS was affected by both categories of online warrants. A higher volume of abuse did not influence attributed VB but did lead to higher PS. This is consistent with definitions of online abuse, such as cyberbullying and cyberstalking, which emphasize frequency of abuse as a key factor which negatively impact victims (e.g., Garett, Lord, & Young, 2016; Menesini & Nocentini, 2009). Frequent abuse manifests as a high volume of abusive content in a chronological record of interactions, for example, as replies to a social media post such as a tweet.

While celebrity tweet valence influenced both measures, its effect on VB was finer grained than its effect on PS. An original negative tweet by the celebrity resulted in greater attributed VB with the incident perceived as less severe. More blame was also attributed following a neutral than positive tweet; however, these two conditions did not differ in PS. Thus, attributed VB was not dependent on the actions of others – celebrities were equally likely to be blamed for an incident, regardless of its severity (i.e., high vs low abuse volume). This supports previous findings, with an adolescent population, that individuals are blamed for cyberbullying if the origin of the abuse can be traced back to something that victims themselves have provoked (DeSmet et al., 2012; Shultz et al., 2014; Weber et al., 2013).

Our findings have implications for the application of Warranting Theory (Walther & Parks, 2002) to online research. Walther et al. (2009) demonstrated that negative content carries more weight than positive information in impression formation. In cases of online abuse, this suggests that observers' impressions would be correspondingly dependent on content valence. While the impact of valence in identity statements (i.e., a celebrity tweet) appears to be continuous (negative tweet = higher blame attribution, positive tweet = lower

blame attribution), the impact of perceptions of severity appear more categorical, with negative content eliciting higher PS than non-negative (i.e., neutral and positive) tweets.

The current results also highlight the importance of identity claims in online impression formation. Warranting Theory (Walther & Parks, 2002) originally posited and early work on Facebook demonstrated (e.g., Scott, 2014; Tong, Van Der Heide, Langwell, & Walther, 2009; Utz, 2010; Walther, Van Der Heide, Kim, Westerman, & Tong, 2008; Walther et al., 2009), that behavioral residue held more weight in impression formation than identity claims. This was explained by the fact that these warrants were viewed as less likely to be overt self-presentation strategies by targets, and thus more legitimate sources of accurate information. More recent work has shown that perceptions of some personality factors can be influenced by identity claims (Fullwood, Quinn, Chen-Wilson, Chedwick, & Reynolds, 2015) and that in some cases these carry more weight than behavioral residue (Scott & Ravenscroft, 2017).

The Warranting Theory of impression formation was originally based on research investigating how observers formed impressions of individuals based on real-world personal space (e.g., college dorm rooms: Gosling, Ko, Mannarelli, & Morris, 2002). Its principles were shown to also apply to online space (e.g., personal web-pages: Vazire & Gosling, 2004) and it has thus formed the basis for many investigations into impression formation online (e.g., Rosenthal-Stott, Dicks, & Fielding, 2015; Scott, Sinclair, Short, & Bruce, 2014). SNSs are evolving to include different types of content/warrants and increasingly complex interactions. The focus of many researchers is thus shifting from traditional personality factors and perceptions of attraction (e.g., Scott, 2014; Scott & Hand, 2016; Walther et al., 2008; Walther et al., 2009) to more diverse personality characteristics and abusive incidents such as in the current experiment (e.g., Scott & Ravenscroft, 2017; Weber et al., 2013). Therefore, it may be pertinent to rethink the application of Warranting Theory in an online

context, perhaps by increasing the categories of warrants (e.g., to distinguish 'third-party comments' from other behavioral residue), or by re-evaluating their impact on distinct measures.

*4.2 Consequences for celebrity social media use*

Our findings that negative-valence tweets both increased attributed celebrity VB and reduced PS of subsequent abusive behavior suggest that celebrities are ascribed responsibility for their online behaviour and that abusive responses to negative content are considered warranted under some circumstances. Such perceptions may be magnified in response to celebrity tweets (rather than similar tweets by public, non-celebrity users) as celebrities' use of social media differs from that of traditional users. Whereas non-celebrities typically utilize SNSs to communicate and share information (Tosun, 2012), celebrities primarily use social media for self-promotion and to gain publicity (Marwick & Boyd, 2010).

Our results demonstrate that, while online activity can be useful for celebrities in ways such as increasing their public exposure (e.g., Gayle & Lawson, 2013), social media use can potentially diminish their brand control and affect their personal wellbeing. Abuse volume (behavioral residue) was shown to affect PS in the current experiment, regardless of original tweet valence. Perceived abuse severity is linked to a variety of negative outcomes in other domains (e.g., Jackson, Gabrielli, Fleming, Tunno, & Makanui, 2014; Neilson, Norris, Bryan, & Stappenbeck, 2017). By engaging in self-promotion on social media, celebrities risk being perceived negatively based on content produced by others, in addition to self-posted content.

The psychological impact of online abuse toward celebrities has not been previously investigated in depth. Anecdotal evidence suggests that affects can be severe with celebrity victims of cyberstalking reporting several negative outcomes associated with their

experiences (e.g., Wykes, 2007). Also, many celebrities have closed social media accounts and ceased activity on platforms as a result of online abuse (Cohen, 2014). These results demonstrate that, even though the impact of abuse may be severe on celebrities, victims are often attributed blame for the abuse perpetrated against them. This could potentially lead to a lack of public sympathy and support which could enhance any negative impact on individuals. Given the serious and lasting negative consequences of online abuse in non-celebrity adult and adolescent populations (e.g., Hinduja & Patchin, 2010), and victims' perceived lack of support (e.g., Crosslin & Golman, 2014), it is important that this issue continue to be investigated. Given the public forum of such abuse against celebrities, there is also the possibility that observers' perceptions of abuse against celebrity victims shapes their perceived online behavioural norms, and thus impressions they form of non-celebrity abuse and victims.

A key finding in the current study was that incidents were perceived as less severe following negative tweets than neutral or positive tweets (which did not differ from each other). This may be due to online norms created as a result of the diverse ways in which Twitter is used by celebrities. While many celebrities use Twitter to express themselves and attempt to connect with their fan base (Marwick & Boyd, 2010), others (e.g., comedians) may use it to generate publicity based on the public's reaction to their tweets (O'Neal, 2014; Quora, 2015). Such individuals may attempt to elicit such reactions by posting 'negative', 'controversial', or 'inflammatory' content, or by posting as an alter ego (e.g., Stephen Colbert's right-wing media persona). In such cases, celebrities who tweet negative content for self-promotion may not only be perceived as 'fair game' for abuse, but this might even be expected by some users. Consequently, in these particular situations, celebrities may remain unaffected by such abuse. Conversely, abuse received in response to a similar tweet which is, in fact, genuine and reflective of a celebrity's true feelings could result in distress and

potentially severe negative consequences (Cohen, 2014). It may be that public perception of this abuse may be skewed due to the incidence of 'solicited' abuse and result in the celebrity receiving little sympathy. Further research which systematically measures the frequency and context of postings (celebrity or otherwise) deemed controversial is required to investigate this possibility more fully.

*4.3 The role of dark triad factors*

The dark triad (DT) of personality has been consistently linked with low levels of empathy (e.g., Doane et al., 2014), suggesting that those high in DT traits may be less likely to perceive abusive incidents from the victim's point of view and appreciate the potential impact such abuse might have on the recipient. In the current study, Machiavellianism, narcissism, and psychopathy were all positively associated with VB in the negative tweet condition. In terms of abuse severity, narcissism and psychopathy were inversely associated with PS in the negative tweet condition. Further, psychopathy was also inversely related to perceived severity in the positive tweet condition. That is, those high in psychopathy perceived abusive tweets as less severe regardless of the original tweet valence, while those high in narcissism only perceived less severity when the original tweet was negative in nature. Multiple regression analysis explored these patterns further and revealed that narcissism was a significant independent predictor of VB in the negative tweet condition, but Machiavellianism and psychopathy were not. Narcissism was also the only significant independent predictor of PS in the negative tweet condition. This suggests that as narcissism increases, VB following negative tweets also increases, while perceived severity of abusive tweets decreases.

This contrasts with findings highlighting Machiavellianism and psychopathy as the DT traits most relevant to cyberbullying engagement and trolling behavior (Buckels et al., 2014; Pabian et al., 2015). One prior study more directly considered narcissism and negative

online behaviors (Fan, Chu, Zhang, & Zhou, 2016), demonstrating that those high in covert (characterized by low, unstable views of self-worth) rather than overt (linked with an inflated self-view) narcissism were more likely to be involved in cyberbullying. This relationship was mediated by self-esteem: low self-esteem underpinned covert narcissism, making individuals more likely to engage in aggressive online behaviors. While differing variants of narcissism and self-esteem were not measured in this study, it is possible that the same underlying mechanisms may explain the current findings. As self-esteem increases, victim-blaming decreases (e.g., Lila, Gracia, & Murgui, 2013), and those with low self-esteem are less likely to defend victims from bullying behavior (Baumeister, Campbell, Krueger, & Vohs, 2003). Future research should explore the role of self-esteem in the links between narcissism, VB, and PS to further clarify these predictions. Additionally, evidence suggests that narcissism is associated with hyperactive threat-monitoring tendencies (Horvath & Morf, 2009). In the current study, it is possible that high levels of narcissism sparked a more intense response towards potentially threatening online content (i.e., the negative tweet), causing participants to blame the celebrity for the subsequent abuse and minimize the severity of what they perceived to be a deserved response.

The present study also found that psychopathy was a significant predictor of PS in the positive tweet condition alone: those high in psychopathy were likely to perceive abuse as less severe when the celebrity tweet was positive in nature. As the tweets used in the positive tweet condition were arguably a reflection of the celebrity's success and happiness (e.g., "*There's only one person whose job is to make you happy. That would be you. Get that right, and all else in life will reflect that happiness*"), it is possible that participants high in psychopathy may believe that celebrities deserved the resulting abuse. A key characteristic of psychopathy is a fundamental belief of superiority over others; those high in the trait often view interactions with others as competitive in nature (Jonason, Wee, & Li, 2015). When

celebrity tweets were positive, those high in psychopathy may have viewed the abusive responses as justified to 'bring the celebrity down'. This is supported by the argument that psychopaths tend to experience envy or contempt for those they perceive to be in a more successful position than themselves (Walker & Jackson, 2017). The predictive link between psychopathy, narcissism and attitudes towards online celebrity abuse suggests that accounting for the characteristics of these traits in anti-trolling campaigns may prove beneficial. For example, as both traits are highly concerned with positive-self presentation, portraying celebrity trolls in a negative or derogatory light may deter this behaviour in those high in these traits.

While narcissism demonstrated predictive value when the celebrity tweet was negative, and psychopathy predicted perceptions of abuse severity when the celebrity tweet was positive, Machiavellianism was not a significant predictor of VB or PS in either condition. Machiavellianism is underpinned by attitudes and behaviors aimed at achieving success at all costs, with little consideration or concern for how one's behavior might impact others (Deluga, 2001). As the central interaction in the present study does not reflect an opportunity for a Machiavellian individual to personally prosper, it is possible that they are indifferent to the incident. Therefore, while Machiavellianism may predict engagement in cyberbullying behavior, which could be used as an advancement tool (Buckels et al., 2014; Pabian et al., 2015), it does not appear to be predictive of attitudes regarding cyberabuse committed by, or expressed towards, others.

*4.4 Limitations and future research*

Despite the novel findings reported here there are some limitations to the research, most notably the gender imbalance in both targets and participants. In this experiment we chose to present exclusively male targets, both because Twitter is a male-dominated platform (Statistia, 2018a) and because male celebrities are most often the targets of abuse on this site

(e.g., Demos, 2014). Although we feel it likely that female celebrities would be perceived in a similar way, further research is required to support this notion.

The majority of participants in the current study were young and female. Although this could mean that the results here do not generalize to society as a whole, previous research looking at victim blame and perceived severity in other areas have typically found that females are less likely than males to attribute blame to victims (e.g., Gerber, Cronin, & Steigman, 2004; Grubb & Turner, 2012), and are more likely to attribute incidents as severe (Ben-David & Schneider, 2005; Davies, Rogers, & Bates, 2008). Thus, effects among male participants may actually be stronger than those reported here.

As well as investigating possible gender differences in victims and participants, future research could examine other aspects of impression formation and perceptions of abuse towards public users compared to celebrities. As celebrities utilize SNSs to build and establish their brand (Marwick & Boyd, 2011; Lim, 2017), any abuse they receive may not only result in increased blame but may also influence how they are viewed by the public (e.g., attractiveness, likeability, trustworthiness). The nature of the abuse could also change observers' impressions. Although Twitter is more established, Snapchat and Instagram are increasingly becoming more popular among celebrities (Lim, 2017). Investigations of observer impressions of abuse which is more picture-based or delivered in response to shared images could test whether the current pattern of findings generalizes to other platforms.

Observers' opinions of online abuse may be more sympathetic to public than celebrity victims, if there are fundamental differences in how and why people use social media (e.g., Page, 2012). Finally, future research may wish to examine the role of other personality factors (e.g., the Big Five; Costa & McCrae, 1985) or individual differences (e.g., self-esteem; Baumeister et al., 2003) in observers' VB and SP in online abuse.

*4.5 Conclusions*

In this study we demonstrated that the content (i.e., valence) of a celebrity's tweet influenced the blame attributed to them by observers following any subsequent abuse they received, and that the volume of abuse influenced both attributed blame and perceived incident severity. Our findings demonstrated the importance of self-generated information (identity claims) in the impressions formed of celebrity victims of online abuse. These findings are significant not only to help us understand how social media users view celebrity victims of cyberabuse, but how these perceptions may help shape online norms. Additionally, DT personality factors influenced observer responses. Higher narcissism led to increased VB and reduced PS after initial negative celebrity tweets, while psychopathy was associated with PS following positive celebrity tweets. Although previous studies have identified an association between DT personality factors and cyberbullying and trolling behavior, we highlight the different factors connected with the interpretation of such events by observers. Taken together, our findings more precisely characterize the potential pitfalls of celebrities utilizing social media for self-promotion.

# References

99Firms (2019). Snapchat Statistics 2019: The facts behind the success. Accessed from:
https://99firms.com/blog/snapchat-statistics/

Alexander, J. (2013) The case of the green vampire: Eco-celebrity, Twitter and youth engagement. *Celebrity Studies*, *4*(3), 353-368. doi: 10.1080/19392397.2013.831625

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211-236. doi: 10.1257/jep31.2.211

Baughman, H.M., Dearing, S., Giammarco, E., & Vernon, P.A. (2012). Relationships between bullying behaviours and the Dark Triad: A study with adults. *Personality and Individual Differences*, 52, 571-575.

Baumeister, R.F., Campbell, J.D., Krueger, J.I., & Vohs, K.D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*(1), 1–44.

Ben-David, S., & Schneider, O. (2005). Rape perceptions, gender role attitudes, and victim-perpetrator acquaintance. *Sex* Roles, *53*(5/6), 385-399. doi: 10.1007/s11199-005-6761-4

Boyd, J. (2019). The most followed accounts on Twitter. Brandwatch. Accessed from:
https://www.brandwatch.com/blog/most-twitter-followers/

Buckels, E.E., Trapnell, P.D., & Paulhus, D.L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97-102.

Bursac, Z., Gauss, C.H., Williams, D.K., & Hosmer, D.W. (2008). Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine*, *3*(17). https://doi.org/10.1186/1751-0473-3-17

Carroll, L.O. (July 2012) Tom Daley Twitter abuse arrest leads to calls to educate people of

legal risks. *The Guardian*., Accessed 19th June 2018 from:

https://www.theguardian.com/technology/2012/jul/31/tom-daley-twitter-abuse-arrest

Cohen, C. (2014). Twitter trolls: The celebrities who've been driven off social media by

abuse. *Daily Telegraph*. Accessed 19th June from:

https://www.telegraph.co.uk/women/womens-life/11238018/Celebrity-Twitter-trolls-

The-famous-people-whove-been-driven-off-social-media-by-abuse.html

Costa, P.T., & McCrae, R.R. (1985) *The NEO personality inventory manual*. Odessa, FL:

Psychological Assessment Resources.

Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative

social potency, and trolling behaviours. *Personality and Individual Differences, 102*,

79-84.

Crosslin, K. & Golman, M. (2014). "Maybe you don't want to face it" – College students'

perspectives on cyberbullying. *Computers in Human Behvaior*, *41*, 14-20. doi:

10.1016/j.chb.2014.09.007

Davies, M., Rogers, P., & Bates, J. (2008). Blame towards male rape victims in a

hypothetical sexual assault as a function of victim sexuality and degree of resistance.

*Journal of Homosexuality*, 55(3), 533-544. doi: 10.1080/00918360802345339

Deluga, R.J. (2001). American presidential Machiavellianism: Implications for charismatic

leadership and rated performance. *The Leadership Quarterly*, *12*, 334-363.

Demos (2014). Male celebrities receive more abuse on Twitter than women. Available at:

https://www.demos.co.uk/press-release/demos-male-celebrities-receive-more-abuse-

on-twitter-than-women-2/

DeSmet, A., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., & De

Bourdeaudhuij, I. (2012). Mobilizing bystanders of cyber-bullying: An exploratory

study into behavioural determinants of defending the victim. *Annual Review of

Cybertherapy and Telemedicine*, 181, 58-63.

Doane, A.N., Pearson, M.R., & Kelley, M.K. (2014). Predictors of cyberbullying perpetration

among college students: An application of the theory of reasoned action. *Computers

in Human Behavior, 36*,154–162. doi: 10.1016/j.chb.2014.03.051

Dooley, J., Pyzalski, J., & Cross, D. (2009). Cyberbullying versus face-to-face bullying: A

theoretical and conceptual review. *Journal of Psychology*, *217*, 182–188.

Douglas, H., Bore, M., & Munro, D. (2012). Construct validity of a two-factor model of

psychopathy. *Psychology, 3*(3), 243-248.

Ellison, N., Steinfeld, C., & Lampe, C. (2007). The benefits of Facebook "friends": Exploring

the relationship between college students' use of online social friends networks and

social capital. *Journal of Computer-Mediated Communication*, *12(4)*, 1143-1168.

Facebook Newsroom (2019) Company Info. Accessed from:

https://newsroom.fb.com/company-info/

Fan, C.Y., Chu, X.W., Zhang, M., & Zhou, Z.K. (2016). Are narcissists more likely to be

involved in cyberbullying? Examining the mediating role of self-esteem. *Journal of

Interpersonal Violence*. https://doi.org/10.1177/0886260516666531.

Fullwood, C., Quinn, S., Chen-Wilson, J., Chadwick. D., & Reynolds, K. (2015). Put on a

smiley face: Textspeak and personality perceptions. *Cyberpsychology, Behavior and

Social Networking. 18*(3), 147-151.

Garde-Hansen, J. & Gorton, K. (2013). Emotion Online: Theorizing affect on the internet.

New York, NY: Palgrave Macmillan.

Gahagan, K., Vaterlaus, J.M., & Frost, L.R. (2016). College student cyberbullying on social

networking sites: Conceptualization, prevalence, and perceived bystander

responsibility. *Computers in Human Behavior, 55(B)*, 1097-1105. doi:

10.1016/j.chb.2015.11.019

Garcia, D., & Sikström, S. (2014). The dark side of Facebook: Semantic representations of

status updates predict the Dark Triad of personality. *Personality and Individual*

*Differences, 67*, 92-96. https://doi.org/10.1016/j.paid.2013.10.001

Garett, R., Lord, L.R., & Young, S.D. (2016). Associations between social media and

cyberbullying: a review of the literature. *MHealth, 2*(46), doi:

10.21037/mhealth.2016.12.01

Gayle, S.S., & Lawson, K. (2013). Twitter as a way for celebrities to communicate with fans:

Implications for the study of parasocial interaction. *North American Journal of*

*Psychology Winter Garden, 15*(2), 339-354.

Gerber, G.L., Cronin, J.M., & Steigman, H.J. (2004). Attributions of blame in sexual assault to

predators and victims of both genders. *Journal of Applied Social Psychology*, *34*(10),

2149-2165.

Gibb, Z.G., & Devereux, P.G. (2014). Who does that anyway? Predictors and personality

correlates of cyberbullying in college. *Computers in Human Behavior*, *38*, 8-16.

Goodboy, A.K., & Martin, M.M. (2015). The personality profile of a cyberbully: Examining

the Dark Triad. *Computers in Human Behavior, 49*, 1-4.

Gosling S.D., Ko, S.J., Mannarelli, T., & Morris, M.E. (2002). A room with a cue:

Personality judgments based on offices and bedrooms. *Journal of Personality and*

*Social Psychology, 82*(3),379-398.

Gray, J. (2003). New audiences, new textualities: Anti-fans and non-fans. *International Journal of Cultural Studies*, *6*(1), 64–81.

Grubb, A., & Turner, E. (2012). Attribution of blame in rape cases: A review of the impact of rape myth acceptance, gender role conformity and substance use on victim blaming. *Aggression and Violent Behavior, 17*, 443-452. 10.1016/j.avb.2012.06.002.

Hermida, A., Lewis, S.C., & Zamith, R. (2014). Sourcing the Arab Spring: A case study of Andy Carvin's sources on Twitter during the Tunisian and Egyptian revolutions. *Journal of Computer-Mediated Communication*, *19*, 479-499. doi: https://doi.org/10.1111/jcc4.12074

Hinduja, S., & Patchin, J. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research*, *14*, 206–221.

Horvath, S., & Morf, C.C. (2009). Narcissistic defensiveness: Hypervigilance and avoidance of worthlessness. *Journal of Experimental Social Psychology*, *45*(6), 1252-1258.

Jackson, Y., Gabrielli, J., Fleming, K., Tunno, A.M., Makanui, P.K. (2014). Untangling the relative contribution of maltreatment severity and frequency to type of behavioral outcome in foster youth. *Child Abuse & Neglect*, *38*(7), 1147-1159. doi: 10.1016/j.chiabu.2014.01.008

James, S., Kavanagh, P.S., Jonason, P.K., Chonody, J.M., & Scrutton, H.E. (2014). The dark triad, schadenfreude, and sensational interests: Dark persnoalities, dark emotions, and dark behaviors. *Persnoality and Individual Differences*, 68, 211-216. https://doi.org/10.1016/j.paid.2014.04.020

Jonason, P.K. & Krause, L. (2013). The emotional deficits associated with the Dark Triad

    traits: Cognitive empathy, affective empathy, and alexithymia. *Personality and*

    *individual Differences*, 55(5), 532-537. https://doi.org/10.1016/j.paid.2013.04.027

Jonason, P.K. & Kroll, C.H. (2015). A Multidimensional View of the Relationship Between

    Empathy and the Dark Triad. *Journal of Individual Differences*, 36, 150-156.

    https://doi.org/10.1027/1614-0001/a000166

Jonason, P.K., Wee, S., & Li, N.P. (2015). Competition, autonomy, and prestige:

    Mechanisms through which the Dark Triad predict job satisfaction. *Personality and*

    *Individual Differences*, *72*, 112-116.

Jones, D.N., & Paulhus, D.L. (2013). Introducing the short dark triad (SD3): A brief measure

    of dark personality traits. *Assessment*, *21(1),* 28-41. doi: 10.1177/1073191113514105

Kircaburun, K., Semetrovics, Z., & Tosuntas, S.B. (2018). Analyzing the Links Between

    Problematic Social Media Use, Dark Triad Traits, and Self-esteem. *International*

    *Journal of Mental Health and Addiction*, 1-12. https://doi.org/10.1007/s11469-018-

    9900-1

Lee, J. & Lim, Y-S. (2016). Generated campaign tweets: The cases of Hillary Clinton and

    Donald Trump. *Public Relations Review*, *42*(5), 849-855.

    https://doi.org/10.1016/j.pubrev.2016.07.004

Lerner, M., & Simmons, C.H. (1966). Observer's reaction to the 'innocent victim':

    Compassion or rejection? *Journal of Personality and Social Psychology*, *4*(2), 203–

    210.

Lila, M., Gracia, E., & Murgui, S. (2013). Psychological adjustment and victim-blaming

    among intimate partner violence offenders: The role of social support and stressful

life events. *The European Journal of Psychology Applied to Legal Context, 5*, 147-153.

Lim, Y.J. (2017). Decision to use either Snapchat or Instagram for most powerful celebrities. *Research Journal of the Institute for Public Relations*, *3*(2), 1016.

Lumsden, K., & Morgan, H. (2017). Media framing of trolling and online abuse: silencing strategies, symbolic violence, and victim blaming. *Feminist Media Studies, 17*(6), 926-940. doi: 10.1080/14680777.2017.1316755

Maple, C., Short, E., Brown, A., Bryden, C., & Salter, M. (2012). Cyberstalking in the UK: Analysis and recommendations. *International Journal of Distributed Systems and Technologies*, *3*(4), 34-51.

Marwick, A. E. & Boyd, D. (2010). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114-133. doi: https://doi.org/10.1177/1461444810365313

Marwick, A. E., & Boyd, D. (2011). To see and be seen: Celebrity practice on Twitter. *Convergence: The International Journal of Research into New Media Technologies*, *17*(2), 139-158. doi: 10.1177/1354856510394539

Mechanic, M.B., Uhlmansiek, M.H., Weaver, T.L., & Resick, P.A. (2000). The impact of severe stalking experienced by acutely battered women: An examination of violence, psychological symptoms, and strategic responding. *Violence and Victims*, *15*, 443–458.

Menesini, E., & Nocentini, A. (2009). Cyberbullying definition and measurement: Some critical considerations. *Journal of Psychology*, *217*(4), 230-232.

Menesini, E., Nocentini, A., Palladino, B.E., Frisén, A., Berne, S., Ortega-Ruiz, R., . . . & Smith, P.K. (2012) Cyberbullying definitions among adolescents: A comparison across six European countries. *Cyberpsychology, Behavior, & Social Networking*, 15(9), 455-463.

Millon, T., & Davis, R.D. (1996). *Disorders of personality: DSM-IV and beyond*. New York: John Wiley & Sons.

Muntean, N., & Petersen, A.H. (2009). Celebrity twitter: Strategies of intrusion and disclosure in the age of technoculture. *M/C Journal: A Journal of Media & Culture*, *12*(5).

Neilson, E.C., Norris, J., Bryan, A.E.B., & Stappenbeck, C.A. (2017). Sexual assault severity and depressive symptoms as longitudinal predictors of the quality of women's sexual experiences. *Journal of Sex & Marital Therapy*, *43*(5), 463–478. http://doi.org/10.1080/0092623X.2016.1208127

O'Neal, S. (2014). Stephen Colbert responds to Twitter satire controversy with more satire on Twitter. AvNews. Accessed 17th August from: https://news.avclub.com/stephen-colbert-responds-to-twitter-satire-controversy-1798267392

Pabian, S., De Backer, C.J.S., & Vandebosch, H. (2015). Dark Triad personality traits and adolescent cyber-aggression. *Personality and Individual Differences*, *75*, 41-46.

Page, R. (2012). The linguistics of self-branding and micro-celebrity on Twitter: The role of hashtags. Discourse & Communication, 6(2), 181-201. https://doi.org/10.1177/1750481312437441

Paulhus, D.L., & Williams, K.M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Psychology, 36*, 556-563.

Quora (August, 2015) How does social media benefit celebrities? Available from: https://www.quora.com/How-does-social-media-benefit-celebrities

Rosenthal-Stott, H.E., Dicks, R.E., & Fielding, L.S. (2015). The valence of self-generated (status updates) and other generated (wall-posts) information determines impression formation on Facebook. *PLoS One*, *10*, e0125064.

Scott, G. G. (2014). More than friends: Popularity on Facebook and its role in impression formation. *Journal of Computer-Mediated Communication, 19*(3), 358-372.

Scott, G.G., & Hand, C.J. (2016). Motivation determines Facebook viewing strategy: An eye movement analysis. *Computers in Human Behavior, 56*, 267-280.

Scott, G.G. & Ravenscroft, K. (2017). Bragging on Facebook: The interaction of content source and focus on online impression formation. *Cyberpsychology, Behavior, & Social Networking, 20*(1), 58-63. doi: 10.1089/cyber.2016.0311

Scott, G.G., Sinclair, J., Short, E., & Bruce, G. (2014). It's not what you say, it's how you say it: Language use on Facebook impacts employability but not attractiveness. *Cyberpsychology, Behavior, and Social Networking, 17*(8), 562-566. https://doi.org/10.1089/cyber.2013.0584

Scott, G.G., Wienercz, S., & Hand, C.J. (2018). The frequency and source of online abuse impacts attribution of victim blame and perceptions of victim attractiveness. *Computers in Human Behavior, 92,* 119-127. https://doi.org/10.1016/j.chb.2018.10.037

Shaver, K.G. (1970). Defensive attribution: effects of severity and relevance on the responsibility assigned for an accident. *Journal of Personality and Social Psychology, 14*, 101-113.

Shultz, E., Heilman, R., & Hart, K.J. (2014). Cyber-bullying: An exploration of bystander behavior and motivation. *Cyberpsychology*: *Journal of Psychosocial Research on Cyberspace, 8*(4), article 3. http://dx.doi.org/10.5817/CP2014-4-3

Short, E., & McMurray, I. (2009). Mobile phone harassment: An exploration of students' perceptions of intrusive texting behaviour. *Human Technology*, *5*(2), 163-180.

Selwyn, N. (2008). A safe haven for misbehaving? An investigation of online misbehavior among university students. *Social Science Computer Review*, *26*(4), 446-465.

Skinner, T. (March, 2018). Rihanna's criticism of Snapchat saw company's value fall by almost $1bn. *NME*. Accessed 19th June from: http://www.nme.com/news/music/rihannas-criticism-snapchat-sees-company-lose-almost-1bn-2266401

Statistia (2018a). Distribution of Twitter users worldwide as of July 2018, by gender. Accessed 17th August 2018 from: https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/

Statistia (2018b). Number of daily Snapchat users from 1st quarter 2014 to 1st quarter 2018 (in millions). Statistia. Accessed 19th June 2018 from: https://www.statista.com/statistics/545967/snapchat-app-dau/

Statistia (2018c). Number of monthly active Instagram users worldwide from January 2013 to September 2017 (in millions). Statistia. Accessed 19th June 2018 from:

https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/

Statistia (2018d). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2018 (in millions). Statistia. Accessed 19th June 2018 from: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). *Predicting Dark Triad personality traits from Twitter usage and a linguist analysis of tweets*. In Paper presented at the International Conference of Machine Learning and Applications (IMCLA).

Thomas, S. (2014). Celebrity in the 'Twitterverse': History, authenticity and the multiplicity of stardom Situating the 'newness' of Twitter. *Celebrity Studies*, *5*(3), 242-255. https://doi.org/10.1080/19392397.2013.845962

Tong, S.T., Van Der Heide, B., Langwell, L., & Walther, J.B. (2008). Too much of a good thing? The relationship between number of friends and interpersonal impressions on Facebook. *Journal of Computer-Mediated Communication*, *13*(3), 531-549.

Tosun, L.P. (2012). Motives for Facebook use and expressing "true self" on the internet. *Computers in Human Behavior*, *28*, 1510-1517.

Utz, S. (2010). Show me your friends and I will tell you what type of person you are: How one's profile, number of friends, and type of friends influence impression formation on social network sites. *Journal of Computer-Mediated Communication*, *15*, 314–335

Van den Bulck, H., Claessens, N., & Bels, A. (2014). 'By working she means tweeting': Online celebrity gossip media and audience readings of celebrity Twitter behavior. *Celebrity Studies*, *5*(4), 514-517. https://doi.org/10.1080/19392397.2014.980655

van Geel, M., Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences, 106*(1), 231-235.

van Geel, M., Vedder, P., Tanilon, J. (2014). Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: A meta-analysis. *JAMA Pediatrics*, *168*(5), 435-442. doi: 10.1001/jamapediatrics.2013.4143.

Vasquez, J. (2018). Kylie Jenner tweeted about snapchat. Then its stock lost $1.3 billion in value. *Time*. Accessed 19th June 2018 from: http://time.com/5170990/kylie-jenner-snapchat-stock-value/

Vazire, S., & Gosling, S.D. (2004). E-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology, 87*, 123- 132

Villlanti, A.C., Johnson, A.L., Ilakkuvan, V., Jacobs, M.A., Graham, A.L., & Rath, J.M. (2017). Social media use and access to digital technology in US young adults in 2016. *Journal of Mediacal Internet Research, 19*(6), e196. doi: 10.2196/jmir.7303

Walker, B.R., & Jackson, C.J. (2017). Moral emotions and corporate psychopathy: A review. *Journal of Business Ethics, 141*(4), 797–810.

Walther, J.B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*(1), 3–43.

Walther, J.B. (1997). Group and interpersonal effects in international computer-mediated collaboration. *Human Communication Research, 23*(3), 342–369.

Walther, J.B., & Parks, M.R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M.L. Knapp & J.A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). Thousand Oaks, CA: Sage.

Walther, J.B., Van Der Heide, B., Hamel, L., & Shulman, H. (2009). Self-generated versus other-generated statements and impressions in computer mediated communication: A test of warranting theory using Facebook. *Communication Research, 36*, 229–253.

Walther, J.B., Van Der Heide, B., Kim, S.Y., Westerman, D., & Tong, S.T. (2008). The role of friends' appearance and behavior on evaluations of individuals on Facebook: Are we known by the company we keep? *Human Communication Research, 34*(1): 28–49.

Watt, N. & McLean, B. (July, 2012). Celebrities and cyberstalkers: The dark side of fame in the internet age. *ABC News*. Accessed 19th June 2018 from: https://abcnews.go.com/Technology/celebrities-cyberstalkers-dark-side-fame-internet-age/story?id=16741230

Weber, M., Ziegele, M., & Schnauber, A. (2013). Blaming the victim: The effects of extraversion and information disclosure on guilt attributions in cyberbullying. *Cyberpsychology, Behavior, and Social Networking, 16*(4), 254-259.

Wu, S., Hofman, J.M., Mason, W.A., & Watts, D.J. (2011). Who says what to whom on Twitter. WWW '1: Proceedings of the 20th international conference on the world wide web. 705-714.

Wykes, M. (2007) Constructing crime: Culture, stalking, celebrity and cyber. *Crime, Media, Culture, 3*, 158-174.

# Appendix A

## Example Experimental Stimulus
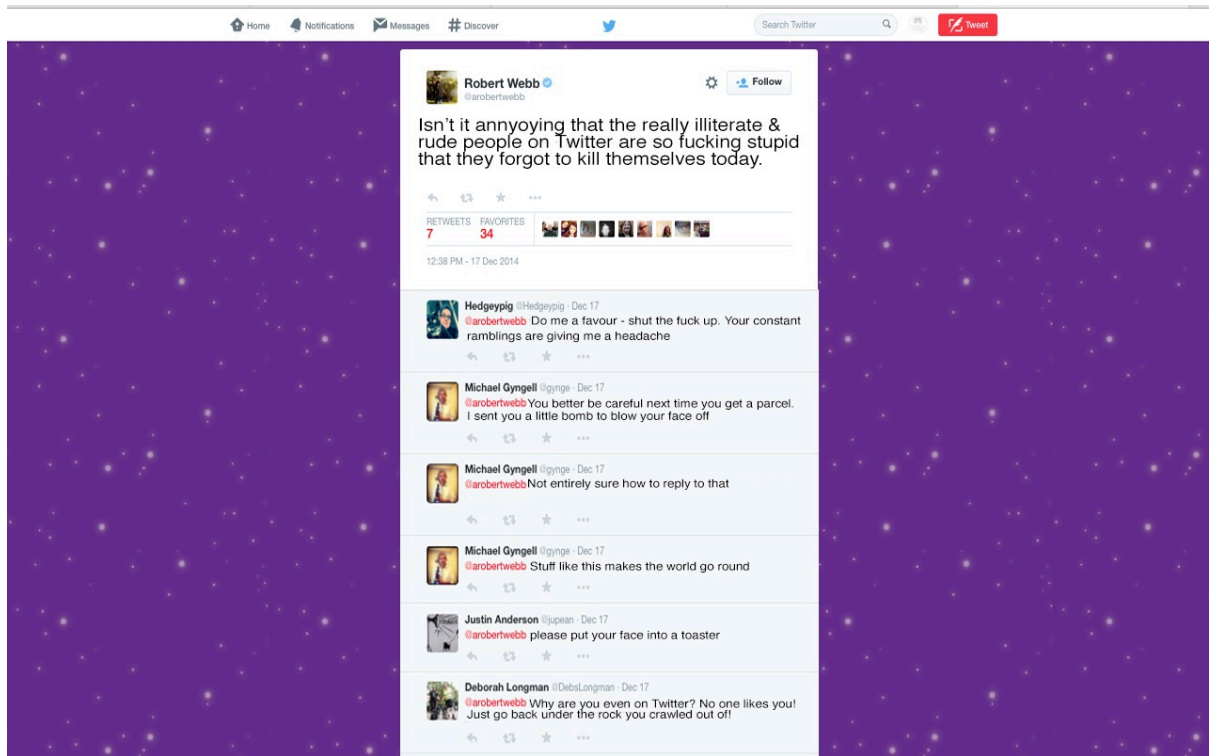
**Positive, Neutral, and Negative Tweets with Valence, Arousal, and Abuse Ratings**

| Positive Tweets | Valence | Arousal | Abuse |
|---|---|---|---|
| Be disciplined about doin' the little things for your goals - daily. Consistency adds up to success. #ChaseYourGreatness | 6.071 | 5.000 | 6.179 |
| Thanks for all your lovely comments. See you tomorrow for big live show! | 6.214 | 4.643 | 6.464 |
| There's only one person whose job is to make you happy. That would be you. Get that right, and all else in life will reflect that happiness | 5.679 | 4.929 | 5.464 |
| We are blessed to have another day to accept the challenge #GoCatchYourDream | 53786 | 4.714 | 6.143 |
| *M* | 53938 | 43821 | 6.063 |
| *SD* | 0.248 | 0.170 | 0.424 |

| Neutral Tweets | Valence | Arousal | Abuse |
|---|---|---|---|
| A Taxi just turned at my house. No idea why. Said he was instructed to take me to Park Lane. Anyone know if I'm supposed to be somewhere?!! | 3.857 | 3.786 | 4.250 |
| Weathers getting chilly. I think summer is over. | 3.824 | 3.071 | 4.750 |
| Aye up peeps. How you diddling? | 4.643 | 3.750 | 5.357 |
| "I'm in the mood to eat chocolate, lay on the sofa and do nothing…. That is all | 4.143 | 3.321 | 4.464 |
| *M* | 4.107 | 3.536 | 4.786 |
| *SD* | 0.380 | 0.346 | 0.480 |

| Negative Tweets | Valence | Arousal | Abuse |
|---|---|---|---|
| Hitler was absolutely right about the Jews. And I don't care how many "white power" redneck followers I lose by Tweeting that | 1.286 | 6.464 | 1.286 |
| You can't get anyone to do anything round here! Bunch of useless fucking c***s! | 1.393 | 5.536 | 1.393 |
| Why do you follow me, you f***ing moron | 1.676 | 4.607 | 1.679 |
| Isn't it annoying that the realy illiterate & rude people on Twitter are so fucking stupid that they forgot to kill themselves today. | 1.536 | 5.500 | 1.357 |
| *M* | 1.473 | 5.527 | 1.429 |
| *SD* | 0.171 | 0.758 | 0.173 |

# Appendix C

## Negative and Neutral Comments with Valence, Arousal, and Abuse Ratings

| Negative Comments | Valence | Arousal | Abuse |
|---|---|---|---|
| Ugly fag | 1.214 | 5.750 | 1.179 |
| Gay fuck | 1.107 | 6.000 | 1.071 |
| I hate you so much. I don't know why I just hate your guts | 1.714 | 4.964 | 1.536 |
| I just love to see what sorta bullshit you manage to bring up with such consistency. Wants me to hate you even more | 1.892 | 4.460 | 2.036 |
| you're a nobody people like you make me sick | 1.321 | 4.571 | 1.607 |
| please put your face into a toaster | 1.857 | 5.107 | 1.536 |
| please do the world a favor, go hang yourself | 1.250 | 5.357 | 1.250 |
| go kill yourself you fucking wanker? | 1.250 | 5.607 | 1.179 |
| wow - does the studitity hurt much? | 2.285 | 4.178 | 2.357 |
| Why are you even on Twitter? No one likes you! Just go back under the rock you crawled out of! | 1.464 | 4.785 | 1.357 |
| Really wishing you're the next Ebola victim… | 1.571 | 6.142 | 1.250 |
| Do me a favour - shut the fuck up! Your constant ramblings are giving me headache! | 1.892 | 4.214 | 1.571 |
| Beginning to wonder what the world is coming to when they let stupid idiots like you on social media | 2.178 | 3.928 | 1.964 |
| Wish you would just find the nearest shark infested waters, cover yourself in shrimp paste and jump in | 2.071 | 5.642 | 1.821 |
| Your so arrogant and self righteous! So full of yourself you arrogant c*** | 1.392 | 5.357 | 1.321 |
| I know where you live! I will find you and kill you! You fucking dick! | 1.285 | 5.750 | 1.107 |
| You better be careful next time you get a parcel, I sent you a little bomb to blow your face off | 1.571 | 5.785 | 1.214 |
| Nice house… shame if someone was to set it on fire while you were in your bed | 1.464 | 5.857 | 1.179 |
| *M* | 1.599 | 5.192 | 1.474 |
| *SD* | 0.357 | 0.691 | 0.364 |

| Neutral Comments | Valence | Arousal | Abuse |
|---|---|---|---|
| winter is coming…. | 4.464 | 4.571 | 5.143 |
| could you wish my mum Suzanne a happy 50th birthday please | 4.607 | 3.750 | 5.393 |
| Would you prefer to be a Wizard or a Jedi? | 4.285 | 3.892 | 4.571 |
| I swear I seen you in costa today! Weren't anywhere near Glasgow were you? | 3.928 | 3.464 | 4.679 |
| Off to the gym I go #WorkingHardOrHardlyWorking | 4.214 | 3.571 | 4.571 |
| Trying to convince my mum to let me have a dog, any chance of some help? | 4.464 | 3.964 | 5.036 |
| Not entirely sure how to reply to that | 3.821 | 3.535 | 4.250 |
| No comment | 3.892 | 3.178 | 4.214 |
| suggestions for tv viewing tonight? | 4.035 | 3.107 | 4.643 |
| Not the best tweet I've seen today, but not the worst | 3.821 | 3.214 | 4.000 |
| tooooo much coffee . . . | 3.857 | 3.357 | 4.143 |
| stuff like this makes the world go round | 4.000 | 3.535 | 4.393 |
| another eclectic tweet | 3.750 | 3.321 | 4.000 |
| I find your attitude confuzing :-/ | 3.571 | 3.500 | 3.929 |
| yesterday go OK? | 4.285 | 3.500 | 4.929 |
| BUT WHAT'S NEXT!? | 4.107 | 4.035 | 4.464 |
| Eating my cereal reading what folks think | 3.892 | 3.285 | 4.393 |
| how'd ya like them apples | 4.107 | 4.035 | 4.214 |
| | | | |
| *M* | 4.061 | 3.601 | 4.498 |
| *SD* | 0.279 | 0.377 | 0.416 |

**Table 1**

**Mean Ratings (SDs) of Valence, Arousal, and Politeness**
**for the Experimental Stimuli of Tweets (Negative, Neutral, Positive)**
**and Comments (Negative, Neutral)**

| Stimulus | Stimulus Valence | Participant Judgment | | |
|---|---|---|---|---|
| | | Valence | Arousal | Politeness |
| Tweet | Negative | 1.46  (0.10) | 5.52  (0.03) | 1.38  (0.03) |
| | Neutral | 4.23  (0.58) | 3.41  (0.18) | 5.05  (0.43) |
| | Positive | 5.88  (0.28) | 4.96  (0.05) | 5.82  (0.51) |
| Comment | Negative | 1.60  (0.36) | 5.19  (0.69) | 1.47  (0.36) |
| | Neutral | 4.06  (0.28) | 3.60  (0.38) | 4.50  (0.42) |

*Note*: Participant judgments were measured on 7-point scales with endpoints 1 and 7 labelled, respectively, as follows:  Valence (very negative – very positive); Arousal (not arousing – very arousing); and Politeness (abusive – polite).

**Table 2**

**Mean Ratings (SDs) of Celebrity Tweets for Victim Blame (VB)**
**and Perceived Severity (PS), with 95% CIs, across Experimental Conditions**

| Tweet Valence | Abuse Volume | VB | VB 95% CI | PS | PS 95%CI |
|---|---|---|---|---|---|
| Negative | Low | 3.71 (1.06) | [3.56-3.87] | 2.88 (1.03) | [2.73-3.03] |
| | High | 3.76 (1.07) | [3.60-3.91] | 3.59 (0.91) | [3.45-3.72] |
| Neutral | Low | 1.67 (0.86) | [1.55-1.80] | 3.33 (0.95) | [3.19-3.47] |
| | High | 1.62 (0.78) | [1.51-1.74] | 4.03 (0.75) | [3.92-4.14] |
| Positive | Low | 1.51 (0.67) | [1.41-1.61] | 3.20 (1.05) | [3.04-3.35] |
| | High | 1.49 (0.67) | [1.39-1.59] | 4.06 (0.82) | [3.94-4.18] |

*Note*:  Participant judgments were measured on 5-point scales with endpoints 1 and 5.

**Table 3**

**Means, SDs, and Pearson's Correlations of Dark Triad Components
and VB and PS Ratings for Positive and Negative Celebrity Tweets**

| | *Mean* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| 1. Psychopathy | 2.09 | 0.59 | 1 | .47** | .46** | .10 | -.15* | .23** | -.25** |
| 2. Narcissism | 2.66 | 0.65 | | 1 | .40** | -.02 | -.04 | .25** | -.27** |
| 3. Machiavellianism | 2.84 | 0.58 | | | 1 | .10 | -.10 | .15* | -.14 |
| 4. Total Positive VB | 3.00 | 1.25 | | | | 1 | -.24** | .11 | -.13 |
| 5. Total Positive Severity | 7.26 | 1.51 | | | | | 1 | -.17* | .47** |
| 6. Total Negative VB | 7.47 | 1.90 | | | | | | 1 | -.41** |
| 7. Total Negative Severity | 6.46 | 1.66 | | | | | | | 1 |

Note. *p<.05; **p<.01. *n*=184.

**Table 4**

**Hierarchical Multiple Regression Analysis, Predicting
Victim Blame (VB) from Negative Tweets
with Psychopathy, Narcissism, and Machiavellianism**

Table 4. *Hierarchical multiple regression analysis predicting VB with psychopathy,
narcissism and Machiavellianism scores (negative tweet condition; n = 184)*

| | $R^2$ | $\Delta R^2$ | $F$ | $B$ | 95% CI |
|---|---|---|---|---|---|
| | .078 | .062 | 5.053** | | |
| Psychopathy | | | | .131 | (-.124, .965) |
| Narcissism | | | | .180* | (.043, .998) |
| Machiavellianism | | | | .021 | (-.466, .603) |

Note. *$p$<.05, **$p$<.01; 95% CI = 95% confidence interval; *n*=184.

**Table 5**

**Hierarchical Multiple Regression Analysis, Predicting**
**Perceived Severity (PS) from Negative Tweets**
**with Psychopathy, Narcissism, and Machiavellianism**

|  | $R^2$ | $\Delta R^2$ | $F$ | $B$ | 95% CI |
|---|---|---|---|---|---|
|  | .093 | .078 | 6.130** |  |  |
| Psychopathy |  |  |  | -.158 | (-.916, .030) |
| Narcissism |  |  |  | -.203* | (-.929, -.099) |
| Machiavellianism |  |  |  | .011 | (-.433, .495) |

Note. *$p$<.05, **$p$<.01; 95% CI = 95% confidence interval; $n$=184.

**Table 6**

**Hierarchical Multiple Regression Analysis, Predicting
Perceived Severity (PS) from Positive Tweets
with Psychopathy**

|  | $R^2$ | $\Delta R^2$ | $F$ | $B$ | $95\%\ CI$ |
|---|---|---|---|---|---|
|  | .021 | .016 | 3.960* |  |  |
| Psychopathy |  |  |  | -.146* | (-.744, -.003) |

Note. *$p$<.05, **$p$<.01; 95% CI = 95% confidence interval; $n$=184.