

Hand, C. J. and Scott, G. G. (2022) Beautiful victims: how the halo of attractiveness impacts judgments of celebrity and lay victims of online abuse. *Computers in Human Behavior*, 130, 107157.

(doi: <u>10.1016/j.chb.2021.107157</u>)

This is the Author Accepted Manuscript.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

https://eprints.gla.ac.uk/262632/

Deposited on: 10 January 2022

Enlighten – Research publications by members of the University of Glasgow <u>http://eprints.gla.ac.uk</u>

Abstract

Research has shown that celebrity and lay victims are attributed blame for cyberabuse incidents. The nature of victim-generated content and abuse volume contribute to victim blame (VB) and perceived severity (PS) of incidents. Complementary cyberbullying research demonstrates that perceived attractiveness co-varies with VB, suggesting a protective 'halo' related to the 'what is beautiful is good' phenomenon. To explore the inter-relationships between victim status (celebrity, lay-user), victim identity claims (initial tweets: negative, neutral, positive), and behavioral residue (abuse volume: low, high), we used a mixed-factors design; victim status was a between-groups factor, whereas initial tweet valence and abuse volume were within-participants factors. We measured perceptions of victim attractiveness, VB, and PS; additionally, we measured participants' (N=309) Dark Triad traits. In general, we found that celebrities received less blame than lay-users, and abuse against celebrities was perceived as more-severe. An exception was when celebrities initially tweeted negative content, in which case they received more blame. VB was influenced by social attractiveness, victim status, and initial tweet valence. PS was determined by abuse volume, task attractiveness, and initial tweet valence. Celebrities appear to be held in higher regard and considered more attractive than other social media users, affording them protection when abused online.

Keywords: attractiveness, cyberabuse, Dark Triad, Twitter, victim blame.

1. Introduction

Online communication, including using social media, has been increasing in society for the past two decades, with many individuals now reliant on digital interactions in both their professional and personal lives (Chaffey, 2019). A consequence of this has been a concomitant increase in online abuse (Allcott & Gentzcow, 2017; Hearn & Hall, 2019), actions which can result in serious negative impact for victims (John et al., 2018). It has been established that both the volume of online abuse, and potentially provocative content posted by the victim themselves, impacts observers' perceptions of both victim blame (VB) and perceived severity (PS) of the incident (e.g., Hand et al., 2021; Scott et al., 2020). This study builds on previous work to further expand our understanding of the relationship between perceptions of VB and incident severity, and the perceived attractiveness of victims, originally proposed by Weber et al. (2013). We compare VB and PS scores, as well as perceived victim attractiveness, from studies in which the volume of abuse and the nature of the 'provoking' content posted by the victim were manipulated, and in which victims were either celebrities or lay-users of social media. We also assess the role played by Dark Triad (Jones & Paulhus, 2013) personality scores of observers. By better understanding how victims of online abuse are viewed, and often blamed, by observers we will be better-placed to minimize the blame attributed to victims for such incidents and mitigate against potential negative fallout.

1.1 Social media use

Over the past 20 years digital communication has become increasingly important, both in the workplace and in individuals' personal lives (Chaffey, 2019). Social media in particular has grown in popularity in this time, particularly with younger people, as more diverse platforms have become available (Allcott & Gentzkow, 2017; Villlanti et al., 2017). Facebook is the largest social media site with 2.85 billion active monthly users, while Twitter is also popular with 340 million users (Aslam, 2021; Tankovsa, 2021). On Twitter, users can broadcast 'tweets' of up to 280 characters and other users can like, comment on, or 'retweet' (share) them. Two ways in which Twitter differs from some other popular social media sites is the extent to which it is utilized heavily by celebrities as well as lay-users (Lee & Lim, 2016; Thomas, 2014), and the fact that often on Twitter users who communicate do *not* know each other offline, in contrast to the majority of Facebook users / 'friends' (Phua et al., 2017).

Social media platforms are perceived as extensions of offline space and users can utilize online cues to form accurate impressions of those they encounter in the online environment (Bak, 2010). The Warranting Theory of online impression formation (Walther & Parks, 2002) posits that impressions are formed using two categories of cues, or warrants: online identity claims (overt claims made by users) and behavioural residue (evidence unintentionally left behind by users), with more weight afforded to behavioural residue, and to negative (vs. positive) information (Walther et al., 2009). Because of the relatively socially-impoverished nature of online (vs. offline) environment, impressions formed using these cues will be slower and more stereotyped (Hyperpersonal Model: Walther, 1996; 1997).

1.2 Online abuse and victim blame

The increase in use of, and communication via, digital technologies has seen a corresponding increase in cases of online abuse (Hearn & Hall, 2019; Méndez et al., 2019; Vakhitova et al., 2019). Such abuse can be comprised of many different acts, and has been classified as trolling, cyberbullying, cyberharassment, and/or cyberstalking (Maple et al., 2012; Menesini & Nocentini, 2009). These categories often overlap, and the constantly evolving nature of technology and online platforms make them difficult to define (Jurgens et al., 2019; Menesini

et al., 2012). One definition of cyberbullying is that of an aggressive act perpetrated via electronic contact (Menesini & Nocentini, 2009), and definitions tend to focus on harmful intent and repetition of actions (Bocij, 2004; Garett et al., 2004; Novo et al., 2014). It is not just layusers who are vulnerable to such online attacks: celebrities commonly utilize social media to promote their brand and interact with fans (Gayle & Lawson, 2013; Page, 2012), but are increasingly targets of online abuse (e.g., Garde-Hansen & Gorton, 2013). Online abuse can have serious negative impacts on victims ranging from depression, anxiety, and loneliness to selfharm and suicide (Gini & Espelage, 2014; Hinduja & Patchin, 2010; John et al., 2018). Online abuse is often more harmful than offline abuse, resulting in enforced changes to personal and work lifestyle in lay users (van Geel et al., 2014). Celebrity social media users are often followed and routinely trolled by 'anti-fans', and such abuse has enforced the closure of celebrity accounts, potentially harming their careers (Cohen, 2014; Gray, 2003). The impact experienced by victims is often confounded by a lack of sympathy and support from both authorities and their personal network, who often underestimate the impact of such deviant acts (e.g., Gahagan et al., 2016; Molluzzo & Lawler, 2012).

In addition to minimizing the perceived severity (PS) of online abuse, observers can also attribute blame to victims for the acts perpetrated against them (e.g., Russell & Hand, 2017). Cyberbullying can be considered acceptable when there is evidence it has been provoked by the victim's initial behaviour (DeSmet et al., 2012), and celebrity victims often do not receive sympathy because they are perceived as using social media as a tool to further their careers (Wu et al., 2011). Theoretical explanations for victim blame (VB) include Belief in a Just World (Lerner & Simmons, 1966) and Defensive Attribution Hypotheses (Shaver, 1970). Just World Theory proposes that people believe the world to be a just place, and that people 'get what they

deserve', so bad things must happen to bad people. Defensive Attribution Hypothesis posits that individuals attempt to increase their own sense of control by attributing an (online) attack to a victim's disposition. Online impressions are formed via a limited number of online warrants (Walther & Parks, 2002), and often exaggerated based on these warrants (Walther, 1996; 1997). It is important to understand how salient cues available on social media may impact PS and VB.

1.3 Online warrants and victim blame

Online VB against both celebrities and lay users of Twitter, as well as PS, has been investigated experimentally in the context of Warranting Theory as both the volume of abuse (behavioural residue) and the initial content posted by the victim (identity claims) were systematically manipulated (Hand et al., 2021; Scott et al, 2020).

Hand et al. (2021) manipulated both the valence of the initial tweet posted by a male victim (negative, neutral, positive) and the volume of abusive comments that were posted in response to said tweet (low, high). They found that VB was highest when the initial tweet was negative, and when the victim received a high volume of abuse. Incidents were perceived to be more severe when they included a high volume of abuse, but PS was unaffected by initial tweet valence. This supported previous findings that cyberbullying victims are often blamed if they can be seen to have provoked any attack (Shultz et al., 2014; Weber et al., 2013).

Scott et al. (2020) employed the same design but with male celebrity Twitter users. Most blame was attributed to celebrities when their initial tweet was negative, and least when their initial tweet was positive; in contrast to perceptions of incidents involving lay-users, a higher volume of abuse led to incidents being perceived as *less* severe. The effect of volume of abuse is in the opposite direction to the finding with lay-users, but consistent with definitions of online abuse which emphasize frequency as a key factor which negatively impacts victims (e.g., Garett et al., 2016).

These results demonstrate that not only do observers rely on online warrants to form impressions of online abuse victims and attribute blame to them, but lay users differ from celebrities in how they are perceived and thus how much blame they receive, and how severe abusive incidents against them are judged to be. Lay-users are attributed blame on the basis of both identity claims and behavioural residue while celebrity victims are blamed only on identity claims, indicating that celebrities may be seen to be using social media for more cynical, selfpromoting purposes than lay-users (Lim, 2017), and are therefore judged more harshly (Hand et al., 2021).

1.4 The role of attractiveness in victim blame

A factor which has long been known to influence impression formation is the physical attractiveness of the individual. The "what is beautiful is good" literature (e.g., Dion et al., 1972) demonstrated that when an individual is physically attractive, a halo effect means they are often assumed to possess positive personality traits. This could be especially important in online environments, where fewer cues are available for impression formation than in real world contexts, and where beauty may be inferred by a perceptually-salient profile image (Hancock & Dunham, 2001; Walther, 1996; 1997). Indeed, the attractiveness halo effect is powerful in the online environment (Bak, 2010) and perceived physical attractiveness of Facebook profile owners increased if they had attractive online friends (Walther et al., 2009).

Weber et al. (2013) investigated the role of attractiveness in online victim blaming and bystander support. They asked participants to provide ratings of perceived VB, PS, and victim attractiveness after reading a cyberbullying scenario and viewing the victim's Facebook profile. The victim was either extraverted or introverted, disclosing either a high or low volume of personal information. Participants attributed a higher level of VB when the victim had disclosed more personal information, but this effect was mediated by perceived attractiveness: higher perceived attractiveness reduced the attributed blame and led to more social support. These results could be explained by a halo effect and highlight the increased culpability and blame that may be experienced by less-attractive victims of online attacks.

1.5 The Dark Triad

In addition to the attractiveness of the actors involved, traits of observers also impact how online abuse is perceived. Personality measures which likely cause individuals to underestimate the severity of online abuse are the Dark Triad personality traits: Machiavellianism, narcissism, and psychopathy (Jones & Paulhus, 2013). All three factors are associated with low empathy (e.g., Doane et al., 2014) and links have been found between the Dark Triad and cyberbullying (e.g., Goodboy & Martin, 2015), with individuals who score highly on these factors more likely to blame victims and to underestimate the impact of online abuse. Individuals who score higher in Machiavellianism are more likely to be manipulative and deceptive in nature, and possess a lack of concern with conventional morality, and a lack of interpersonal affect (Deluga, 2001). Narcissism is characterized by high levels of vanity and self-enhancement tendencies, as well as a sense of superiority, entitlement, and selfishness (Paulhus & Williams, 2002). If an individual scores highly in psychopathy they may exhibit aversive interpersonal (e.g., callousness, remorselessness) and behavioural (e.g., anti-social behaviour, impulsivity) characteristics (Douglas et al., 2012).

For observations of online abuse against lay-users, psychopathy predicted PS following positive or negative initial tweets, whereas Machiavellianism predicted PS following neutral

tweets (Hand et al., 2021). When observing abuse against celebrities, narcissism predicted both VB and PS following negative tweets, and psychopathy predicted PS following positive tweets (Scott et al., 2020). Observers scoring high in the Dark Triad are less-able to perceive online abuse from victims' perspective or understand the negative impact that such actions may cause.

1.6 The current study

The current study expands upon the previous studies by Hand et al. (2021) and Scott et al. (2020). Those studies focused primarily on the interplay between the effects of volume and source of abuse on perceptions of VB and PS, in addition to the role of observer dark traits. We expand this work by examining the inter-relationships between cyberabuse, VB, PS, and perceptions of victim attractiveness; attractiveness is both theoretically and empirically important in relation to concepts of VB and PS (e.g., Dion et al., 1972; Weber et al., 2013). We not only assess the role played by attractiveness in attributions of VB and PS, but also compare differences across lay and celebrity victims of online abuse, as celebrity victims are conceptualized differently from lay users (Lim, 2017).

Weber et al. (2013) found that the perceived attractiveness of the victim mediated VB when the initial manipulation was an identity claim: the amount of personal information disclosed. We extend that work by including an initial manipulation that included both behavioural residue (volume of abuse) and an identity claim directly relevant to the abuse (initial tweet valence). Furthermore, we used McCroskey and McCain's (1974) model of attractiveness, which in addition to physical attractiveness (i.e., beauty) includes social attractiveness (i.e., desirability as a friend) and task attractiveness (i.e., desirability as a work colleague); finally, we included participant-observers' Dark Triad personality factors, which have been previously found to influence perceptions of VB and PS (Hand et al., 2021; Scott et al., 2020). Celebrities are held in high esteem in society and are often idolised by fans and the general public (Brooks, 2018). Their higher status, success in their domain, and popularity are all traits which have been related to perceived attractiveness (Dion et al., 1972). We therefore predicted:

*H*₁: Celebrity victims would be rated as more attractive than lay victims

Perceived attractiveness has previously been shown to relate to attributed victim blame (Weber et al., 2013). Celebrity victims are predicted to be perceived as more attractive than lay victims, therefore we predicted:

> H_{2a} : Celebrity victims would be attributed less blame for the abuse they receive on Twitter than lay victims.

*H*_{2b}: *Ratings of perceived attractiveness would be correlated with attributed victim blame*

Given previous theories of online impression formation (Warranting Theory: Walther & Parks, 2002), we predicted that the identity claims and behavioural residue manipulated in our experimental design would impact perceptions of the victims and the abusive incident, in line with previous findings (Hand et al., 2021; Scott et al., 2020). Specifically, we predicted:

 H_{3a} : Victims would be attributed more blame if they have initially authored a negative (vs. neutral or positive) tweet.

 H_{3b} : Incidents would be perceived as more severe when the victim receives a high (vs. low) volume of abuse.

2. Method

2.1 Design

A 2 (victim status: celebrity, layperson) \times 3 (initial tweet valence: negative, neutral, positive) \times 2 (abuse volume: low, high) mixed factor quasi-experimental design was developed,

using an online survey method. The victim status factor was established as a between-subjects factor so that other aspects of the stimuli (initial tweet, responses) could be held identical and rotated across conditions (but without having to repeat stimuli within-participants). Initial tweet valence and abuse volume were repeated-measures factors. We measured direct victim blame, perceived incident severity, and perceptions of victims' social-, physical-, and task-attractiveness as covariates; we additionally measured observers' Dark Triad characteristics. An a priori power analysis was conducted using G*Power 3.1.9.2, with a smallest anticipated effect size of *f*=.20, an α =.05, and desired power =0.95 (Cohen, 1988); this analysis suggested a minimum required sample of 168 participants.

2.2 Participants

A total of 309 participants completed one of the two surveys (230 females, 77 males, 2 non-binary participants). Participants were aged between 16 and 53 years (M_{age} =23.60 years, SD_{age} =5.18). For each victim-type survey, participant recruitment utilised a mixture of oncampus poster advertisement and social media posts disseminated via the authors' networks; all data collection took place online. Participants were offered no financial incentive(s) for taking part. A summary of participant group demographics across the between-subjects factor of victim status is presented in Table 1.

Table 1. Participant demographics by celebrity and layperson victim surveys

			Sex			A	ge	
Victim	N	Female	Male	Non-Binary	Min.	Max.	Mean	SD
Celebrity	184	146	38	0	17	50	22.61	5.11
Layperson	125	84	39	2	16	53	25.06	4.94

In both sub-samples there is an imbalance towards over-representation of female participants. The imbalance is more pronounced in the celebrity victim sub-sample [$\chi^2(1)$ =4.795, p=.029; n.b., this analysis considers the relationship between sample sub-group females and males only]¹. The difference between the ages of the sub-samples was statistically significant [Mann-Whitney U test: z=7.43, p<.001]; however, in reality, it is debatable whether there is a real difference between these samples – the shape of the distribution of ages across each group was similar, the skew was similar across groups (celebrity=2.57, layperson=2.62), and psychologically / developmentally, it is unlikely that these samples are different from one another in 'age'.

2.3 Materials and measures

Profile owners ('victims') were either celebrities or lay-persons. Celebrity targets were identical to those of Scott et al. (2020) and lay-persons were identical to those of Hand et al. (2021). These papers provide full details of victim identity selection and control of extraneous variables (such as profile picture attractiveness) for celebrities and lay-persons, respectively. Norming with a separate sample ensured that celebrities were familiar to participants to the point of being recognizable, but they did not elicit any strong positive or negative feelings. Furthermore, tweets and comments were controlled on scales of valence (positive-negative), arousal (arousing-not arousing), and politeness (polite-abusive). Full details of stimulus creation, pre-test norming, valence and abusiveness ratings are available within Scott et al. (2020) and summarised by Hand et al. (2021).

Participants were presented with six stimuli representing artificial Twitter interactions. Each stimulus consisted of an initial victim tweet followed by six replies from lay-persons unconnected to our participant-observers. Initial tweets were either tonally negative, neutral, or positive, and replies were either socially acceptable or abusive – to manipulate the abuse volume factor, 2 abusive + 4 neutral replies constituted a low volume of abuse, whereas 4 abusive + 2 neutral replies reflected a high volume of abuse. The content of victim tweets and replies were identical to those used by Scott et al. (2020).

The measurement of direct victim blame (DVB) and perceived incident severity (PS) was based on the work of Weber et al. (2013) and in-line with Scott et al. (2020) and Hand et al. (2021). DVB was established via four items (e.g., "*Did the victim provoke the abuse*?") with responses made on a five-point Likert-type scale (1 = *strongly disagree* – 5 = *strongly agree*; current data set Cronbach's α =.922, *F*(3,5559)=65.284, *p*<.001). PS (e.g., "*How severe was the abuse*?") was assessed via two items with the same five-point Likert-Type scale as DVB (current data set Cronbach's α =.691, *F*(1,1854)=758.222, *p*<.001).

Perceived social-, physical-, and task-attractiveness (McCroskey & McCain, 1974) were each measured on five-item, seven-point Likert-type scales ($1 = strongly \ disagree - 7 = strongly \ agree$). Participants disclosed their agreement with statements such as "*I think he could be a* friend of mine." [social attractiveness; current data set Cronbach's α =.897, F(4,7580)=50.344, p<.001], "*I find him very attractive physically*" [physical attractiveness; current data set Cronbach's α =.848, F(4,7580)=699.188, p<.001] or "*I couldn't get anything accomplished with him.*" [task attractiveness; current data set Cronbach's α =.871, F(4,7580)=60.228, p<.001].

Dark Triad scores were obtained using the SD3 (Jones & Paulhus, 2013), composed of 27 items – nine items per sub-dimension. Likert-type responses were required, on a five-point scale (1= strongly disagree, 5= strongly agree). All measures were found to be reliable: Machiavellianism (e.g., "You should wait for the right time to get back at people."; current data set Cronbach's α =.758, *F*(8,2496)=111.631, *p*<.001); narcissism ("People see me as a natural") *leader.*"; current data set Cronbach's α =.772, *F*(8,2496)=32.175, *p*<.001); psychopathy (e.g., *"Payback needs to be quick and nasty."*; Cronbach's α =.727, *F*(8,2496)=42.911, *p*<.001).

2.4 Procedure

Prior to data collection, ethical approval was granted by the host university's Ethics Committee, following British Psychological Society protocols (BPS, 2014). Data were gathered online using the QuestionPro platform (https://www.questionpro.com/). Participants were given full information and instructions prior to giving informed consent. Participants provided brief demographic details (age, gender-sex, national identity). Participants were then presented with one of six versions of our survey – six versions were used to counterbalance the order or the tweet stimuli and rotate tweet content across valence / volume conditions. Participants read initial tweets and replies before providing ratings of DVB and PS via the instruments described in section 2.3. After reading and rating all victim tweets and replies, participants completed the Dark Triad items (see section 2.3). To conclude, participants were provided with a short debrief and were thanked for their participation. Participation lasted approximately 20 minutes, on average.

2.5 Data analysis

Data were pre-screened for missing values at item-level, and the 309 participants' data that was retained represented those who provided complete responses. Independent and combined effects of victim status, initial tweet valence, and abuse volume on DVB and PS were analysed through two three-way mixed factor analyses of variance (ANOVAs). For comparisons related to the main effect(s) of valence, Bonferroni corrections were applied. Relationships between DVB, PS and Dark Triad variables were assessed via Pearson's correlations (onetailed). Finally, linear regression analyses evaluated the relationships between ratings of DVB and PS as outcomes, with victim status, initial tweet valence, volume of abuse, perceived attractiveness (social, physical, task), and observer Dark Triad dimensions as potential predictors. Pre-checks suggested that the assumptions relevant to each of the above analytical approaches were generally supported; in any cases where there was a violation, adjustments and alternative approaches are detailed in section 3.

3. Results

3.1 ANOVAs

3.1.1 Direct Victim Blame (DVB) and Perceived Severity (PS)

Mean ratings and 95% confidence intervals for DVB and PS across conditions are

presented in Table 1.

Tweet Valence	Abuse Volume	Celebrity DVB (SD)	Layperson DVB (SD)	Celebrity PS (SD)	Layperson PS (SD)
Negative	Low	14.85 (0.27)	13.40 (0.33)	5.76 (0.15)	5.69 (0.19)
	High	15.03 (0.27)	14.18 (0.33)	7.17 (0.13)	7.05 (0.16)
Neutral	Low	6.69 (0.23)	9.64 (0.28)	6.65 (0.15)	6.01 (0.18)
	High	6.48 (0.22)	10.47 (0.27)	8.07 (0.14)	6.89 (0.16)
Positive	Low	6.04 (0.21)	9.45 (0.25)	6.39 (0.15)	5.85 (0.19)
	High	5.97 (0.21)	10.00 (0.25)	8.13 (0.14)	7.14 (0.17)
Tweet Valence	Abuse Volume	Celebrity DVB 95% CI	Layperson DVB 95% CI	Celebrity PS 95% CI	Layperson PS 95% CI
Negative	Low	[14.31, 15.39]	[12.74, 14.05]	[5.45, 6.06]	[5.20, 5.94]
	High	[14.49, 15.56]	[13.53, 14.83]	[6.91, 7.44]	[6.73, 7.36]
Neutral	Low	[6.24, 7.14]	[9.10, 10.18]	[6.36, 6.95]	[5.65, 6.36]
	High	[6.05, 6.92]	[9.94, 11.00]	[7.80, 8.33]	[6.57, 7.21]
Positive	Low	[5.64, 6.45]	[8.95, 9.94]	[6.09, 6.69]	[5.48, 6.21]

Table 1. Mean Ratings (plus standard deviations) of Direct Victim Blame (DVB) and Perceived Severity (PS) with 95% Confidence Intervals across conditions.

Note: Participant judgments were measured on five-point scales with (lower scores = less blame). Figures are rounded to 2DP.

For DVB, analysis revealed a small-to-medium significant main effect of victim status on DVB [F(1,307)=75.934, p<.001, $\eta_p^2=.198$]; laypersons (11.19) faced greater DVB than celebrities (9.18). Analysis of PS data revealed a small, significant effect of victim status [F(1,307)=15.330, p<.001, $\eta_p^2=.048$]; incidents involving laypersons were perceived as less-severe (6.42) than those involving celebrities (7.03).

Analysis revealed a large and significant main effect of initial tweet valence on DVB $[F(2,614)=602.019, p<.001; \eta_p^2=.662]$. Planned follow-up comparisons demonstrated that greater DVB was attributed in relation to negative initial tweets (14.36) than neutral (8.32; p<.001) or positive tweets (7.87, p<.001); furthermore, DVB was greater in following a neutral tweet than a positive tweet (p=.001). A small, significant main effect of initial tweet valence on PS was observed [$F(2,614)=18.515, p<.001; \eta_p^2=.057$]. Planned follow-up comparisons demonstrated that PS of abuse generated in relation to negative initial tweets (6.39) was less than that associated with neutral (6.90; p<.001) or positive tweets (6.88, p<.001); PS of abuse related to neutral and positive victim tweets was perceived equivalently (p>.999).

The main effect of abuse volume on DVB was small but significant [F(1,307)=10.743, p=.001; $\eta_p^2=.034$]; DVB was greater following a high volume of abuse (10.36) than a low volume of abuse (10.01). A large main effect of abuse volume on PS was found [F(1,307)=304.534, p<.001; $\eta_p^2=.034$]; PS was greater following a high volume of abuse (7.41) than a low volume of abuse (6.04).

A significant victim status × initial tweet valence interaction was observed on DVB ratings [F(2,614)=86.086, p<.001, η_p^2 =.219]. Further analysis revealed that when initial tweets

were negative, DVB attributed to celebrities (14.94) was greater than that attributed to laypersons (13.79; p=.003); when initial tweets were neutral, DVB attributed to celebrities (6.59) was less than that attributed to laypersons (10.05; p<.001); when initial tweets were positive, DVB attributed to celebrities (6.01) was less than that attributed to laypersons (9.73; p<.001).

For PS ratings, a significant victim status × initial tweet valence interaction was observed $[F(2,614)=8.773, p<.001, \eta_p^2=.028]$. Further analysis revealed that when initial tweets were negative, incidents involving celebrities (6.47) were seen as equally serious to those involving laypersons (6.31; p=.410); when initial tweets were neutral, PS involving celebrity victims (7.36) was greater than layperson victims (6.45; p<.001); when initial tweets were positive, celebrity incidents (7.26) were seen as more severe than that layperson incidents (6.49; p<.001).

The victim status × abuse volume interaction was also found to be significant for DVB $[F(1,307)=12.997, p<.001, \eta_p^2=.041]$. When abuse volume was low, DVB attributed to celebrities (9.19) was lower than that attributed to laypersons (10.83; p<.001); when abuse volume was high, DVB attributed to celebrities (9.16) was less than that attributed to laypersons (11.55; p<.001). Alternatively, there was a non-significant simple main effect of abuse volume for celebrity DVB (p=.797), whereas DVB of laypersons was greater following a high vs. a low volume of abuse (p<.001). For PS, the victim status × abuse volume interaction was non-significant [F(1,307)=3.798, p=.052]

There was no evidence of an interaction between the effects of initial tweet valence and abuse volume on DVB ratings [F < 1] or PS ratings [F(2,614)=2.923, p=.055]. There was no evidence of a three-way interaction between victim status, initial tweet valence, and abuse volume on DVB [F < 1] or PS [F(2,614)=1.981, p=.139].

3.1.2 Attractiveness

Descriptive statistics for social, physical, and task attractiveness are presented in Table 2.

Table 2. Mean ratings (plus standard deviations) of social attractiveness (SA), physical attractiveness (PA), and task attractiveness (TA) with 95% Confidence Intervals across conditions.

Tweet Valence	Abuse Volume	Celebrity SA (SD)	Layperson SA (SD)	Celebrity PA (SD)	Layperson PA (SD)	Celebrity TA (SD)	Layperson TA (SD)
Negative	Low	15.53 (0.43)	15.40 (0.52)	16.65 (0.39)	14.94 (0.48)	15.66 (0.42)	13.37 (0.35)
	High	15.22 (0.44)	13.88 (0.53)	17.33 (0.40)	14.22 (0.48)	16.17 (0.45)	13.33 (0.37)
Neutral	Low	23.42 (0.40)	23.17 (0.49)	21.46 (0.40)	18.24 (0.49)	18.52 (0.45)	16.21 (0.37)
	High	23.94 (0.44)	21.94 (0.53)	21.66 (0.37)	18.33 (0.45)	18.64 (0.47)	15.60 (0.38)
Positive	Low	24.21 (0.38)	23.94 (0.46)	20.58 (0.36)	19.37 (0.44)	22.62 (0.54)	19.94 (0.44)
	High	24.14 (0.41)	23.54 (0.50)	21.28 (0.36)	18.60 (0.44)	23.30 (0.51)	21.17 (0.42)
Tweet Valence	Abuse Volume	Celebrity SA 95% CI	Layperson SA 95% CI	Celebrity PA 95% CI	Layperson PA 95% CI	Celebrity TA 95% CI	Layperson TA 95% CI
Negative	Low	[14.68, 16.38]	[14.37, 16.43]	[15.87, 17.43]	[14.00, 15.89]	[14.83, 16.50]	[12.68, 14.06]
	High	[14.36, 16.08]	[12.84, 14.93]	[16.54, 18.11]	[13.27, 15.18]	[15.28, 17.06]	[12.59, 14.06]
Neutral	Low	[22.62, 24.21]	[22.20, 24.13]	[20.67, 22.24]	[17.29, 19.19]	[17.64, 19.40]	[15.48, 16.93]
	High	[23.07, 24.80]	[22.90, 22.99]	[20.94, 22.38]	[17.45, 19.20]	[17.72, 19.56]	[14.85, 16.36]
Positive	Low	[23.47, 24.95]	[23.04, 24.84]	[19.87, 21.29]	[18.47, 20.20]	[21.56, 23.68]	[19.06, 20.81]
	High	[23.33, 24.95]	[22.55, 24.52]	[20.57, 21.98]	[17.74, 19.46]	[22.30, 24.29]	[20.35, 22.00]

Note: Figures are rounded to 2DP.

Victim status had a small significant effect on perceived social attractiveness [F(1,307)=4.918,

p=.027, $\eta_p^2=.016$]; celebrities were perceived as more socially-attractive (21.07) than laypersons

(20.31). Victim status had a significant, small effect on perceived physical attractiveness

[F(1,307)=52.291, p<.001, $\eta_p^2=.146$]; celebrities were perceived as more physically-attractive

(19.82) than laypersons (17.28). Analysis revealed a significant effect of victim status on task

attractiveness ratings [F(1,307)=63.307, p<.001, η_p^2 =.171]; participants rated celebrities (19.15) higher in task attractiveness than laypersons (16.60).

A large, significant main effect of initial tweet valence on social attractiveness was found [F(2,614)=348.681, p<.001; η_p^2 =.532]. Planned follow-up comparisons demonstrated that those who posted negative initial tweets were perceived as less socially-attractive (15.01) than those posting neutral (23.12; p < .001) or positive tweets (23.96, p < .001); those posting positive tweets were perceived as more socially-attractive than those posting neutral content (p=.007). A medium-sized, significant main effect of initial tweet valence on physical attractiveness was found [F(2,614)=115.064, p<.001; $\eta_p^2=.273$]. Planned follow-up comparisons demonstrated that those who posted negative initial tweets were perceived as less physically-attractive (15.79) than those posting neutral (19.92; p < .001) or positive tweets (19.95, p < .001); those posting positive tweets and neutral content were perceived as equally physically attractive (p>.999). A mediumto-large, significant main effect of initial tweet valence on task attractiveness was observed $[F(2,614)=203.082, p<.001; \eta_p^2=.398]$. Planned follow-up comparisons demonstrated that those who posted negative initial tweets were perceived as less task-attractive (14.63) than those posting neutral (17.24; $p \le .001$) or positive tweets (21.76, $p \le .001$); those posting positive tweets were rated as more task-attractive than those posting neutral tweets (p < .001).

A small main effect of abuse volume on perceived social attractiveness was found $[F(1,307)=6.213, p=.013; \eta_p^2=.020]$; those receiving a high volume of abuse were seen as less socially-attractive (20.44) than those only receiving a low volume of abuse (20.94). There was no effect of abuse volume on perceived physical attractiveness [F<1], and no significant effect of abuse volume on perceived task attractiveness [F(1,307)=3.097, p=.079].

There was no evidence of a victim status × initial tweet valence interaction on social attractiveness [F<1] or physical attractiveness [F(1,307)=2.233, p=.108].

Analysis revealed a significant victim status × abuse volume interaction on social attractiveness ratings [F(1,307)=7.335, p=.007]. When celebrities were the targets of abuse, abuse volume had no impact on social attractiveness (p=.865), whereas when laypersons were targeted, those laypersons receiving a high volume of abuse were perceived as less socially-attractive (20.84) than laypersons receiving lower volumes of abuse (19.79; p=.001). A significant victim status × abuse volume interaction was also found for ratings of physical attractiveness [F(1,307)=6.689, p=.010, $\eta_p^2=.021$]. Laypersons were perceived as less physically-attractive (17.51) than celebrities (19.56) at a low volume of abuse (p<.001), and this gap widens when abuse volume is high (laypersons: 17.05, celebrities: 20.09; p<.001). There was no evidence of an interaction between victim status and abuse volume on perceived task attractiveness [F<1].

There was no significant evidence of an interaction between initial tweet valence and abuse volume on social attractiveness [F(2,614)=1.034, p=.356] or physical attractiveness [F<1]. However, there was a significant initial tweet valence × abuse volume interaction on task attractiveness ratings $[F(2,614)=3.383, p=.035, \eta_p^2=.011]$. When initial tweets were negative, there was no difference in perceived task attractiveness between those receiving low or high volumes of abuse (p=.411); similarly, when initial tweets were neutral, there was no simple main effect of abuse volume (p=.460). However, following positive tweets, those who received a high volume of abuse (21.28) were perceived as less task-attractive than those who received only a low volume of abuse (22.24; p=.008). There was no significant evidence of a three-way interaction between victim status, initial tweet valence, and abuse volume on social attractiveness [F(2,614)=1.010, p=.365], or physical attractiveness [F(2,614)=1.455, p=.234], or task attractiveness [F<1].

3.2 Correlations

Correlational analyses are summarised in Table 3.

	1	2	3	4	5	6	7	8
1. Machiavellianism		.405	.447	062	143	.365	.311	246
		***	***		**	***	***	***
2. Narcissism			.466	.106	.164	.132	.161	155
			***	*	**	*	**	**
3. Psychopathy				020	005	.301	.261	287
						***	***	***
4. Social Attractiveness					.412	.531	249	.111
					***	***	***	*
5. Physical Attractiveness						.381	123	.047
						***	*	
6. Task Attractiveness							417	.194
							***	***
7. DVB								270

8. PS								

 Table 3. Pearson's correlation coefficients (one-tailed)

Note: All *dfs* = 309. Blank cells represent *p*>.05. **p*<.05, ***p*<.01, ****p*<.001

As can be seen in Table 3, participants who scored more highly in Dark Triad characteristics (Machiavellianism, narcissism, psychopathy) were more likely to blame victims and to perceive incidents as less severe. Furthermore, victims who were perceived as more attractive (social, physical, task) were less likely to be blamed for their abuse (conversely, less attractive victims were more likely to be victim-blamed). Incidents involving victims who were more socially- and task-attractive were perceived as more-severe, whereas there was no linear relationship between physical attractiveness and PS.

3.3 Regression analyses

Building on these correlational results, we considered the relationships between DVB and PS as outcomes, as predicted by victim status (dummy coded with 'celebrity' as baseline), initial tweet valence (with 'negative' as baseline), abuse volume ('low' as baseline), victims' perceived social-, physical-, and task-attractiveness, and observer Dark Triad characteristics as predictors. Two separate analyses were performed, each using a stepwise entry method with an *F*-probability for entry of .05 (and .10 for removal). Assumptions were met.

3.3.1 Direct Victim Blame (DVB)

The model returned yielded an R=.746, with an adjusted R^2 =.555 [F(7,1846)=330.956, p<.001]. Model co-efficients are summarised in Table 4.

	Unstandardized		Sta	andardize	95% CI for B		
Predictor	В	Std. Error	β	t	р	Lower	Upper
SA	-0.132	0.016	-0.193	8.251	<.001	-0.163	-0.101
Valence:Pos	-5.093	0.218	-0.506	23.345	<.001	-5.521	-4.665
Valence:Neu	-4.967	0.209	-0.493	23.754	<.001	-5.378	-4.557
Status	1.417	0.160	0.147	8.839	<.001	1.103	1.732
ТА	-0.097	0.017	-0.126	5.592	<.001	-0.131	-0.063
Psychopathy	0.448	0.137	0.057	3.268	.001	0.717	0.787
Machiavellianism	0.399	0.140	0.051	2.848	.004	0.673	0.738

Table 4. Model co-efficients – DVB

Note: SA = social attractiveness; Pos = positive; Neu = neutral; TA = task attractiveness.

As can be seen in Table 4, increased perceived social attractiveness is reliably associated with a decrease in DVB, and increases in initial tweet valence from negative-to-neutral and negative-to-positive are associated with marked decreases in DVB. The 'cost' of 'not being a celebrity' victim can be seen in increased victim-blaming (conversely, there is a 'benefit' to being a celebrity victim through decreased victim-blaming). Task attractiveness increases are associated with decreased DVB, whereas increases in observer psychopathy and Machiavellianism are each associated with increased DVB. Collectively, this model explains approximately 56% of variability in DVB judgements.

3.3.2 Perceived Severity (PS)

The model returned yielded an R=.425, with an adjusted R^2 =.177 [F(7,1846)=57.974, p<.001]. Model co-efficients are summarised in Table 5.

	Unsta	ndardized	Sta	undardize	95% CI for B		
Predictor	В	Std. Error	ß	t	р	Lower	Upper
Volume	1.406	0.089	0.332	15.740	<.001	1.231	1.581
Psychopathy	-0.466	0.083	-0.133	5.620	<.001	-0.628	-0.303
ТА	0.028	0.009	0.081	3.264	.001	0.011	0.045
Status	-0.361	0.097	-0.084	3.713	<.001	-0.552	-0.170
Valence:Neu	0.462	0.116	0.103	3.985	<.001	0.235	0.690
Valence:Pos	0.348	0.125	0.077	2.777	.006	0.102	0.593
Machiavellianism	-0.211	0.085	-0.061	2.488	.013	-0.377	-0.045

Table 5. Model co-efficients – PS	Table	5.	Model	co-efficients –	PS
--	-------	----	-------	-----------------	----

Note: SA = social attractiveness; Pos = positive; Neu = neutral; TA = task attractiveness.

A greater volume of abuse is associated with a greater PS. As observer psychopathy increases, PS goes down. As perceptions of victim's task attractiveness increases, as does PS.

The 'cost' of 'not being a celebrity' victim can be seen in decreased perceived severity (vice versa, there is a 'benefit' to being a celebrity victim through heightened perceived severity). Increases in initial tweet valence from negative-to-neutral and negative-to-positive are associated with greater PS. Increases in Machiavellianism are associated with decreased PS. Collectively, this model explains approximately 18% of variability in PS ratings.

4. Discussion

The aim of this study was to expand the findings of previous research into the blame attributed by observers to celebrity and lay victims of online abuse, and the perceived severity of the incidents, by examining the role played by victim attractiveness. We re-analysed data from previous studies which measured VB and PS on Twitter following manipulation of the initial tweet valence (negative, neutral, or positive) and volume of abuse (low, high) including observers' perception of victims' social-, physical-, and task-attractiveness. The roles played by the Dark Triad personality factors of participants were also analysed.

4.1 The cause célèbre

Celebrities received less blame than lay victims, and abuse against them was perceived as more severe. This varied by the valence of the initial tweet and the volume of abuse received. Following negative tweets, more blame was attributed to celebrities than lay-victims, but following neutral or positive tweets, celebrity victims were blamed less and incidents were perceived as more severe (partially supporting H_{sa}). While celebrity victim blame did not change with the volume of abuse, lay users were blamed more when they received a high (vs. low) volume of abuse, and they were blamed more than celebrity victims in both volume conditions.

Celebrities seem to hold a privileged position on social media whereby they are held in higher esteem than other users, are less responsible for abuse they receive, and such abuse is regarded as being more serious. This could be due in part to a tacit acknowledgment of celebrity status, particularly in Western society where celebrity is idolised (Brooks, 2018). Celebrities utilize social media in a different way to lay users: lay users primarily use the site to maintain existing relationships and friendships, to share information such as photos with others, and to organise activities (Garcia & Sikstrom, 2014; Tosun, 2012), whereas celebrities use social media to promote themselves and causes they support and establish an online brand (Alexander, 2013; Lim, 2017; Page, 2012), and to interact with fans (Marwick & Boyd, 2011). With this difference in status and varying patterns of use it appears that celebrity users are thought of differently to lay users. The results show that observers are aware of this and that celebrities will be judged harshly if they take advantage of, or abuse, their position – when victims made an initial negative tweet the pattern was reversed, with celebrities attracting more blame.

4.2 Attractiveness

While initial tweet valence and volume of abuse had an impact on perceived attractiveness, overall, celebrity victims were rated as higher in social-, physical-, and taskattractiveness than lay victims (supporting H₁). All victims who made an initial negative tweet were viewed as lower in social-, physical-, and task-attractiveness, while all victims who received a high volume of abuse were rated as lower in social attractiveness. While abuse volume did not impact perceptions of celebrity social attractiveness, lay victims were rated as being lower in social attractiveness following a high volume of abuse. Lay victims were perceived as being less physically attractive than celebrity victims, with the gap higher when a low volume of abuse was received.

This reinforces our assertion that celebrities are seen not just as different, but as *better* than lay users on social media, and that this advantage applies to all types of attractiveness.

23

There is also evidence that these perceptions of attractiveness shielded celebrity users from some of the negative judgements experienced by lay victims, namely reduced perceptions of social attractiveness following a high volume of abuse.

These results are in line with the 'what is beautiful is good' literature (Dion et al., 1972) in which individuals who are perceived as high in physical attractiveness are assumed to also possess other desirable traits, while conversely those low in physical attractiveness are assumed to possess undesirable traits ('ugly is bad'; McKelvie & Coley, 1993). This applies specifically to characteristics related to social competence and interpersonal ease including: altruism, warmth, popularity, kindness, and sincerity (Bassili, 1981; Eagly et al., 1991; Walther et al., 2009). In both original studies (Hand et al., 2021; Scott et al., 2020) the attractiveness of victims was controlled in norming studies, but results show that celebrities were rated higher in physical attractiveness, as well as social- and task-attractiveness by observers. These mirroring effects – that celebrities are regarded as more attractive than lay users, and attributed less blame – are likely related and linked to the higher esteem in which celebrity users are held online.

4.3 Victim blaming

In order to more clearly delineate the role played by the different types of attractiveness in observer judgments of online abuse we built regression models to determine the predictors of attributed VB and perceived incident severity. VB was lower when perceived social attractiveness was high (supporting H_{2b}) and when the victim was a celebrity (vs. lay user), but VB increased when the initial tweet was negative (vs. neutral or positive; supporting H_{3a}), and when task attractiveness was low. Incidents were perceived as more severe when the volume of abuse was high (supporting H_{3b}), and when perceived task attractiveness was high. Incident severity was also judged to be lower following initial negative tweets.

The results of the regression models suggest that social attractiveness is the main driver for observer judgments of VB, while task attractiveness is relevant to attributed blame and PS. This represents an expansion to Weber et al.'s (2013) original findings of an association between victim attractiveness and blame. They used a single item 5-point scale where participants were asked to rate the scenario 'victim' Sophia as attractive to unattractive, with the specific type of attractiveness ambiguous. Although it has been assumed that the principal category of attractiveness is physical, or beauty (Dion et al., 1972), it may be that in online environments other types of attractiveness are either more important or more salient. In online environments fewer cues are available than in the real world, so impressions are formed quickly based on relatively little information (Hancock & Dunham, 2001; Walther, 1996; 1997). It may be the case that with a profile picture as the only salient indicator of physical attractiveness, observers rely more on judgments of social and task attractiveness to form impressions. Depending on the social media platform interactions in the form of published tweets and comments may provide more complex information about a profile owner's character. This may be especially true with text-rich online interactions like the stimuli used by Scott et al. (2020) and Hand et al. (2021).

Perceived social attractiveness was found to be a driver of victim blame by Scott et al. (2019) in a study which manipulated volume and source of abuse on Facebook. This study investigated perceptions of abuse against lay-users of Facebook and manipulated both the volume of abuse (low vs. high) and the source of the abusive comments (single author vs. multiple author). Social attractiveness was rated lowest, and VB highest, in instances where the same source communicated a low volume of abuse. Results were explained by the existence of 'dark' friendships on Facebook where a victim is perceived as being popular, and the abuse intended to be humorous, when there was a high volume of abuse. Although the dynamics of

friendship and communication differ between Facebook and Twitter, with Twitter users less likely to have established offline friendships (Scott et al., 2019), the link between social attractiveness and victim blame appears to extend across platforms and suggests that physical beauty may not be as important online as it is in the real world.

In addition to the roles played by celebrity status, initial tweet valence, and volume of abuse, Psychopathy and Machiavellianism both negatively correlated to both VB and PS. Both traits have previously been related to cyberbullying behaviour (Buckels et al., 2014; Pabian et al., 2015). These Dark Triad traits (as well as narcissism) are associated with low empathy levels and a reduced ability to take others' perspectives (e.g., Doane et al., 2014) with individuals high in them unable to relate to the viewpoint of victims and understand the negative impact of abuse. Individuals high in psychopathy think of themselves as superior to others and are highly competitive in nature (Jonason et al., 2015). Such individuals may view abuse as bringing others down, and therefore increasing their own position by comparison. For celebrities, who may be considered to occupy an elevated position to start with, this could be viewed as 'taking them down' (Walker & Jackson, 2017), whereas lay victims may be viewed as especially inferior, and so any impact of abuse on them will likely be minimized (Scott et al., 2020). Individuals high in Machiavellianism are focused on achieving success without any concern for how their actions might impact others (Deluga, 2001). For this reason, they might be less sympathetic to any victim of abuse, and not be as sensitive to the severity of abusive incidents.

4.4 Future Directions

Future research should focus on further investigating the role played by different types of attractiveness, and different levels of celebrity, or categories of social media user. Task attractiveness, for example, was shown to contribute to perceptions of both VB and PS. The role

of task attractiveness could increase if online abuse occurs specifically in the context of the workplace, or in a school setting / university academic discipline. Additionally, the current results demonstrate an 'advantage' held by celebrities over lay users in terms of public sympathy when they receive abuse (celebrities are blamed less and incidents against them are judged to be less severe). Future work should examine 'levels' of celebrity to see if there is a linear relationship between fame and sympathy, or if celebrities are categorised differently depending on the reason for their fame. Also, it may be possible that some lay users can achieve celebrity-like status online, and thus gain a certain amount of immunity from online abuse, e.g., if they are in a position of authority in society such as a teacher. Future work could also focus on differentiating the effects of 'celebrity' and 'familiarity' in observer judgments. Even though celebrities are conceptualised differently to lay users on social media, observers would have a level of familiarity with them, whereas in the current study all lay-user victims were unfamiliar to participants. It would be interesting to compare judgments of abuse towards celebrities and 'familiar', known lay users.

We have discussed the online environment as being distinct from the offline world in terms of the importance of social attractiveness relative to physical attractiveness, but there is often an overlap between on- and off-line communities (Scott et al., 2019), and abuse often spills over from the real world to the online domain (Modecki et al., 2014). More research is required to differentiate the roles played by different types of attractiveness to judgments of victims, and of abusive incidents, which transcend cyberspace. Furthermore, we need to understand different victim demographics profiles – for example, any potential differences when victims are prototypically female, when victims of colour are abused, and of course the intersectionality of such variables.

There exists a need to further establish the generalizability of these findings across platform. We have discussed the theory that in experiments focusing on Twitter (Hand et al., 2021; Scott et al., 2020) and Facebook (Scott et al., 2019) social attractiveness may be more important than physical attractiveness as the environment is text-heavy and social attractiveness may be more perceptually salient. Other social media platforms such as Snapchat, Instagram, and TikTok are becoming increasingly popular, particularly among celebrities (Lim, 2017). As these sites are more image- and video-based than the older Twitter and Facebook, physical attractiveness may play more of a role in observer judgments.

4.5 Conclusion

In conclusion, this study aimed to further knowledge about perceptions of online abuse by examining the role played by attractiveness in observer judgments of victim blame and perceive incident severity. We analysed data from the studies by Scott et al. (2020) and Hand et al. (2021), who manipulated the volume of abuse as well as the valence of an initial tweet by either celebrity or lay victims, also taking into account perceptions of social-, physical-, and taskattractiveness. Victims were blamed less when they were celebrities, when the initial tweet was not-negative, and when they were perceived as socially attractive but low in task attractiveness. Incidents were perceived as more severe when victims received a high volume of abuse, when they were perceived as being high in task attractiveness, and when the initial tweet was notnegative. Results highlight the fact that in instances of online abuse, high perceived social attractiveness is associated with less attributed blame. This also suggests that in online environments social, rather than physical, attractiveness may be more salient and therefore a more important factor in observer judgments.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Both authors contributed equally to all facets of this this work.

Funding

None.

Acknowledgments

None.

¹Footnote

We considered that the imbalance of female-to-male participants could potentially influence our results. In order to address this, we conducted additional analyses including Participant Sex as a factor. The findings from these analyses can be found in **Supplementary Material A**. In short, our main results are not the result of participant sex / the participant sex imbalance.

References

- Alexander, J. (2013). The case of the green vampire: Eco-celebrity, Twitter and youth engagement. *Celebrity Studies*, *4*(3), 353–368. https://doi.org/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. <u>https://doi.org/10.1257/jep31.2.211</u>
- Aslam, S. (2021). Twitter by the numbers: Stats, demographics & fun facts. Retrieved May 24, 2021 from: <u>https://www.omnicoreagency.com/twitter-statistics/</u>
- Bak, P.M. (2010). Sex differences in the attractiveness halo effect in the online dating environment. *Journal of Business and Media Psychology*, 1, 1-7. <u>https://journalbmp.de/wp-content/uploads/JBMP-1-2010-Halo-Bak.pdf</u>
- Bassili, J. N. (1981). The Attractiveness Stereotype: Goodness or Glamour?, Basic and *Applied* Social Psychology, 2(4), 235-252, <u>https://doi.org/10.1207/s15324834basp0204_1</u>
- Bocij, P. (2004). *Cyberstalking: Harassment in the Internet Age and how to Protect your Family.* Westport, CT: Praeger Publishers.
- British Psychological Society (2014). Code of Human Research Ethics. Leicester: BPS.
- Brooks, S.K. (2018). FANatics: Systematic literature review of factors associated with celebrity worship, and suggested directions for future research. *Current Psychology*, 40, 864–886. <u>https://doi.org/10.1007/s12144-018-9978-4</u>.
- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality* and Individual Differences, 67, 97–102. <u>https://doi.org/10.1016/j.paid.2014.01.016</u>

- Chaffey, D. (2019). *Global social media research summary 2020*. Smartinsights.com. Accessed 31/3/2020 at: <u>https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/</u>
- Cohen, C. (2014). Twitter trolls: The celebrities who've been driven off social media by abuse.
 Daily Telegraph. Accessed 19/06/2019 from: <u>https://www.telegraph.co.uk/women/womens-life/11238018/Celebrity-Twitter-trolls-</u> <u>The-famous-people-whove-been-driven-off-social-media-by-abuse.html</u>
- Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Routledge.
- Deluga, R.J. (2001). American presidential Machiavellianism: Implications for charismatic leadership and rated performance. *The Leadership Quarterly*, *12*, 334–363. <u>https://doi.org/10.1016/S1048-9843(01)00082-0</u>
- 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.
- Dion, K.K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285-290. <u>https://doi.org/10.1037/h0033731</u>
- 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. <u>https://doi.org/10.1016/j.chb.2014.03.051</u>
- Douglas, H., Bore, M., & Munro, D. (2012). Construct validity of a two-factor model of psychopathy. *Psychology*, *3*(3), 243–248. <u>https://doi.org/10.4236/psych.2012.33034</u>

- Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*(1), 109–128. <u>https://doi.org/10.1037/0033-</u> 2909.110.1.109
- 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.

https://doi.org/10.1016/j.chb.2015.11.019

- Garcia, D., & Sikstrom, 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. <u>https://doi.org/10.1016/j.paid.2013.10.001</u>.
- Garde-Hansen, J. & Gorton, K. (2013). *Emotion Online: Theorizing affect on the internet*. New York, NY: Palgrave Macmillan.
- Garett, R., Lord, L.R., & Young, S.D. (2016). Associations between social media and cyberbullying: A review of the literature. *MHealth*, 2(46),

https://doi.org/10.21037/mhealth.2016.12.01

- Gini, G., & Espelage, D.L. (2014). Peer victimization, cyberbullying, and suicide risk in children and adolescents. *JAMA 312*, 545. <u>https://doi.org/10.1001/jama.2014.3212</u>
- Goodboy, A.K., & Martin, M.M. (2015). The personality profile of a cyberbully: Examining the Dark Triad. *Computers in Human Behavior*, 49, 1–4. <u>https://doi.org/10.1016/j.chb.2015.02.052</u>
- Hancock, J.T., & Dunham, P.J. (2001). Impression formation in computer-mediated communication revisited: An analysis of the breadth and intensity of impressions.

Communication Research, 28(3), 325–347.

https://doi.org/10.1177/009365001028003004

- Hand, C.J., Scott, G.G., Brodie, Z.P., Xilei, Y., & Sereno, S.C. (2021). Tweet valence, volume of abuse, and observers' Dark Tetrad personality factors influence victim-blaming and the perceived severity of Twitter cyberabuse. *Computers in Human Behavior Reports,* 100056. <u>https://doi.org/10.1016/j.chbr.2021.100056</u>
- Hearn, J., & Hall, M. (2019). New technologies, image distribution and cyberabuse. *NOTA News*, 88, 10–13.
- Hinduja, S., & Patchin, J. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research*, *14*, 206–221. <u>https://doi.org/10.1080/13811118.2010.494133</u>
- Gray, J. (2003). New Audiences, New Textualities: Anti-Fans and Non-Fans. International Journal of Cultural Studies, 6(1), 64–81.

https://doi.org/10.1177/1367877903006001004

- John, A., Glendenning, A.C., Marchant, A., Montgomery, P., Stewart, A., Wood, S., Lloyd, K., & Hawton, K. (2018). Self-harm, suicidal behaviours, and cyberbullying in children and young people: Systematic review. *J Med Internet Res*; 20(4):e129. <u>https://doi.org/10.2196/jmir.9044</u>
- 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. <u>https://doi.org/10.1016/j.paid.2014.08.026</u>
- 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. <u>https://doi.org/10.1177/1073191113514105</u>

- Jurgens, D., Hemphill, L., & Chandrasekharan, E. (2019). A just and comprehensive strategy for using NLP to address online abuse. In *Proceedings of ACL* <u>https://arxiv.org/abs/1906.01738</u>
- 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. <u>https://doi.org/10.1037/h0023562</u>
- 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.
- 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. (2011). To see and be seen: Celebrity practice on Twitter. Convergence. *The International Journal of Research into New Media Technologies*, 17(2), 139–158. <u>https://doi.org/10.1177/1354856510394539</u>
- McCroskey, J.C., & McCain, T.A. (1974). Measurement of interpersonal attraction. *Speech Monographs, 41*, 261–266. <u>https://doi.org/10.1080/03637757409375845</u>
- McKelvie, S. J., & Coley, J. (1993). Effects of crime seriousness and offender facial attractiveness on recommended treatment. *Social Behavior and Personality: An International Journal*, 21(4), 265–277. <u>https://doi.org/10.2224/sbp.1993.21.4.265</u>

- 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.
 https://doi.org/10.1089/cyber.2012.0040
- Méndez, I., Jorquera, A.B., Ruiz-Esteban, C., Martínez-Ramón, J.P., & Fernández-Sogorb, A. (2019). Emotional intelligence, bullying, and cyberbullying in adolescents.
 International Journal of Environmental Research and Public Health, 16(23), 4837.
 https://doi.org/10.3390/ijerph16234837
- Modecki, K.L., Minchin, J., Harbaugh, A.G., Guerra, N.G., & Runions, K.C. (2014). Bullying Prevalence Across Contexts: A Meta-analysis Measuring Cyber and Traditional Bullying, *Journal of Adolescent Health*, 55(5), 602-611.
 https://doi.org/10.1016/j.jadohealth.2014.06.007
- Molluzzo, J.C., & Lawler, J. (2012). A study of the perceptions of college students on cyberbullying. *Information Systems Education Journal*, 10(4), 84-109. <u>https://isedj.org/2012-10/N4/ISEDJv10n4p84.html</u>
- Novo, F., Pereira, F., & Matos, M. (2014). Cyber-aggression among Portuguese adolescents: A study on perpetration, victim offender overlap and parental supervision. *International Journal of Cyber Criminology*, 8(2), 94-110.
- 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. <u>https://doi.org/10.1016/j.paid.2014.11.015</u>

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

- Paulhus, D.L., & Williams, K.M. (2002). The Dark Triad of personality: Narcissism,
 Machiavellianism, and psychopathy. *Journal of Research in Psychology*, *36*, 556–563.
 https://doi.org/10.1016/S0092-6566(02)00505-6
- Phua, J., Jin, S.V., & Kim, J. (2017). Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat. *Computers in Human Behavior*, 72, 115–122.

https://doi.org/10.1016/j.chb.2017.02.041

- Russell, K.J., & Hand, C.J. (2017). Rape myth acceptance, victim blame attribution and Just World Beliefs: A rapid evidence assessment. *Aggression and Violent Behavior*, 37, 153–160. <u>https://doi.org/10.1016/j.avb.2017.10.008</u>
- Scott, G.G., Brodie, Z.P., Wilson, M.J., Ivory, L., Hand, C.J., & 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*, 109–119. <u>https://doi.org/10.1016/j.chb.2019.09.020</u>
- 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. <u>https://doi.org/10.1016/j.chb.201 8.10.037</u>
- 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. <u>https://doi.org/10.1037/h0028777</u>

- 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. <u>http://dx.doi.org/10.5817/CP2014-4-3</u>
- Stever, G.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, 15(2), 339–354. <u>https://psycnet.apa.org/record/2014-07390-009</u>
- Tankovsa, H. (2021). Facebook: Number of active monthly users worldwide 2008-2021. Retrieved May 24, 2021 from: <u>https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/</u>
- 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. <u>https://doi.org/10.1080/19392397.2013.845962</u>
- Tosun, L. P. (2012). Motives for Facebook use and expressing "true self" on the internet. *Computers in Human Behavior*, 28, 1510–1517.

https://doi.org/10.1016/j.chb.2012.03.018

- Vakhitova, Z., Alston-Knox, C.L., Reynald, D.M., Townsley, M.K., & Webster, J.L. (2019).
 Lifestyles and routine activities: Do they enable different types of cyber abuse?
 Computers in Human Behavior, 101, 225–237 <u>https://doi.org/10.1016/j.chb.2019.07.012</u>
- 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. <u>https://doi.org/10.1001/jamapediatrics.2013.4143</u>

- Walker, B. R., & Jackson, C. J. (2017). Moral emotions and corporate psychopathy: A review.
 Journal of Business Ethics, 141(4), 797–810. <u>https://doi.org/10.1007/s10551-016-3038-5</u>
- Walther, J.B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1), 3–43. https://doi.org/10.1177/009365096023001001
- Walther, J.B. (1997). Group and interpersonal effects in international computer-mediated collaboration. *Human Communication Research*, 23(3), 342–369. <u>https://doi.org/10.1111/j.1468-2958.1997.tb00400.x</u>
- 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. https://doi.org/10.1177/0093650208330251
- 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.
 https://doi.org/10.1089/cyber.2012.0328
- 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. <u>https://doi.org/10.1145/1963405.1963504</u>

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 Medical Internet Research, 19(6), e196. <u>https://doi.org/10.2196/jmir.7303</u>

Supplementary Material A

We considered the possibility that the imbalance of the gender-sex composition between our celebrity and lay survey participants may have influenced our results. To resolve these concerns, we conducted an additional four-way mixed-factor analysis with the addition of participant Sex to the effects of initial tweet Valence, abuse Volume, and victim Status. Please note that this analysis only included participants who identified as female or male, as there were too few non-binary participants to constitute an additional level of this variable.

Direct Victim Blame

The main effect of participant sex on DVB was non-significant [F(1,303)=3.732, p=.054]. The Valence × Sex interaction was non-significant [F(1,303)=2.332, p=.128]. The Sex × Status interaction was non-significant [F<1]. The Volume × Sex interaction was significant [F(1,303)=6.290, p=.013, $\eta_p^2=.020$]. Follow-up comparisons revealed that the simple main effect of Sex was significant when abuse volume was low (p=.005) with males demonstrating greater VB (10.64) than females (9.83). When abuse volume was high, there was no simple main effect of participant Sex (p=.473).

The Valence × Volume × Sex interaction was non-significant [F<1]. The Volume × Status × Sex interaction was non-significant [F<1]. The Valence × Status × Sex interaction was significant [F(1,303)=7.620, p=.001]. The pattern of follow-up comparisons can be summarized thusly: when victims were celebrities and initial tweets were negative, males (15.96) attributed greater VB than females (14.67; p=.031). When victims were laypersons and initial tweets were positive, males (10.59) attributed more VB than females (9.37; p=.015). The simple simple main effect of participant sex was non-significant for all other comparisons (all ps>.087). The Valence × Volume × Status × Sex interaction was non-significant [F(1,303)=1.604, p=.202].

Perceived Incident Severity

The main effect of participant sex on PS was non-significant [F(1,303)=2.242, p=.135]. The Valence × Sex interaction was non-significant [F<1]. The Sex × Status interaction was nonsignificant [F<1]. The Volume × Sex interaction was significant [F(1,303)=5.061, p=.025, $\eta_p^2=.016$]. Follow-up comparisons revealed that the simple main effect of Sex on PS was nonsignificant when abuse volume was low (p=.754). When abuse volume was high, there was a simple main effect of participant Sex (p=.014), with females (7.52) rating PS as higher than males (7.05).

The Valence × Volume × Sex interaction was non-significant [F(1,303)=1.610, p=.201]. The Volume × Status × Sex interaction was non-significant [F<1]. The Valence × Status × Sex interaction was non-significant [F<1].

The Valence × Volume × Status × Sex interaction was non-significant [F(1,303)=1.604, p=.202].

Social Attractiveness

The main effect of participant sex on social attractiveness was non-significant [F<1]. The Sex × Status interaction was non-significant [F<1]. The Volume × Sex interaction was non-significant [F<1]. The Valence × Sex interaction was significant [F(1,303)=8.090, p<.001, η_p^2 =.020]. Follow-up comparisons revealed that the simple main effect of Sex was significant only when initial tweets were negative, with males (16.56) perceiving victims as more socially-

attractive than females (14.48; p=.001). When initial tweets were neutral, there was no simple main effect of Sex (p=.204), nor was there a simple main effect of Sex on social attractiveness when initial tweets were positive (p=.137).

The Valence × Volume × Sex interaction was non-significant [F<1]. The Volume × Status × Sex interaction was non-significant [F(1,303)=1.902, p=.169]. The Valence × Status × Sex interaction was non-significant [F(1,303)=2.439, p=.088].

The Valence × Volume × Status × Sex interaction was significant [F(1,303)=3.516, p=.030; η_p^2 =.011]. The pattern of comparisons can be summarized thusly. When victims were celebrities, there were no simple simple simple main effects of Sex on social attractiveness ratings (all *p*s>.058). When victims were lay-persons, and when initial tweets were negative, males (17.69) perceived victims as more socially-attractive than females (14.55) when abuse volume was low (*p*=.005), and males (16.26) perceived victims as more social-attractive than females (13.01) when abuse volume was high (*p*=.004). Additionally, when initial tweets were positive, initial tweets were positive, and abuse volume was high, males (22.00) perceived layperson victims as less socially-attractive than females (24.23; *p*=.040).

Physical Attractiveness

The main effect of participant sex on physical attractiveness was non-significant [F(1,303)=2.828, p=.094].

The Valence × Sex interaction was non-significant [F<1]. The Volume × Sex interaction was non-significant [F<1]. The Sex × Status interaction was significant [F(1,303)=5.668, p=.018; η_p^2 =.018]. Follow-up comparisons revealed that the simple main effect of Sex on physical attractiveness was significant when victims were celebrities (p=.003), with females

(21.21) rating celebrities as more attractive males (19.97). When victims were lay-persons, there was no simple main effect of participant Sex (p=.632).

The Valence × Volume × Sex interaction was non-significant [F<1]. The Volume × Status × Sex interaction was non-significant [F<1]. The Valence × Status × Sex interaction was non-significant [F(1,303)=1.151, p=.317].

The Valence \times Volume \times Status \times Sex interaction was non-significant [F<1].

Task Attractiveness

The main effect of participant sex on task attractiveness was significant [F(1,303)=4.850, p=.028; η_p^2 =.016]. Females (20.35) rated victims as more task-attractive than males (19.87).

The Volume × Sex interaction was non-significant [F<1]. The Valence × Sex interaction was significant [F(1,303)=9.335, p<.001; η_p^2 =.030]. Follow-up comparisons revealed that the simple main effect of Sex on task attractiveness was non-significant when initial tweets were negative (p=.071). The simple main effect of Sex was significant when initial tweets were neutral (p=.049), with females (20.69) rating victims as more task-attractive than males (20.09). The simple main effect of Sex was significant when initial tweets were positive (p<.001), with females (21.92) rating victims as more task-attractive than males (20.46).

The Sex × Status interaction was significant [F(1,303)=17.806, p<.001; η_p^2 =.056]. Follow-up comparisons revealed that the simple main effect of Sex on task attractiveness was non-significant when victims were celebrities (p=.142); when victims were lay-persons, there was a simple main effect of participant Sex (p<.001), with females (21.31) rating lay-victims as more task-attractive than males (19.94). The Valence × Volume × Sex interaction was non-significant [F(1,303)=2.158, p=.116]. The Volume × Status × Sex interaction was non-significant [F<1]. The Valence × Status × Sex interaction was significant [F(1,303)=9.748, p<.001; η_p^2 =.031]. The pattern of effects can be summarized thusly. When victims were celebrities, there were no simple simple main effects of participant Sex on task attractiveness. When victims were lay-persons, there was no simple simple main effect of Sex when initial tweet were negative (p=.074). When tweets were neutral, females (21.99) perceived lay-victims as more task-attractive than males (20.22; p<.001). When tweets were positive, females (25.14) perceived lay-victims as more task-attractive than males (21.85; p<.001).

The Valence × Volume × Status × Sex interaction was non-significant [F(1,303)=1.142, p=.320].