



Narcissism and Cyber Trolling: The Roles of Self-Esteem, Digital Empathy, and Online Anonymity

Tayyeb Ramazan¹, Muhammad Farrukh², Raza Zaidi³, Anam Amjad⁴

¹Lecturer, School of Creative Arts, University of Lahore.

²Assistant Professor, School of Creative Arts, University of Lahore.

³Associate Professor, School of Creative Arts, University of Lahore.

⁴Student, University of Lahore.

DOI: <https://doi.org/10.71145/rjsp.v3i4.533>

Abstract

The emergence of digital communication platforms has enhanced the possibilities of social interaction as well as the ease of disruptive activities like cyber trolling. This paper investigated the connection between narcissism and cyber trolling behaviour, the mediating effect of self-esteem and empathy, and the moderating effect of perceived online anonymity. The data samples were gathered from online users using a structured questionnaire using validated measures of narcissism, cyber trolling, self-esteem, empathy, and the perceived anonymity of the internet under the quantitative cross-sectional study design. The analysis of data was performed with the aid of descriptive statistics, reliability analysis, Pearson correlation, regression analysis, and Hayes PROCESS macro to test the mediation and moderated mediation modelling. The results indicated that narcissism had a strong positive predictability to cyber trolling behaviour. This relationship was partially mediated by self-esteem and empathy, meaning that the narcissistic traits mediate trolling behaviour both in the self-evaluative and emotional sense. Cyber trolling was negatively related to empathy and positively related to self-esteem, which underscores the protective quality of empathy as compared to self-esteem. The perceived anonymity on the internet was independent of cyber trolling and did not mediate the direct or indirect relationships between narcissism and cyber trolling or these two factors acting through self-esteem and empathy.

Keywords: Narcissism, Cyber Trolling, Self-Esteem, Empathy, Perceived Online Anonymity

Introduction

The prevalence of digital communication technologies has altered the nature of social interaction, allowing people to communicate, share their views, and form identities online. Online forums, social networking sites, and interactive media platforms have become the centre of social life, allowing individuals to be more connected and express themselves more freely (Valkenburg et al., 2022). Simultaneously, these media have presented additional challenges regarding online behaviour, such as the introduction of disruptive and antagonistic behaviour, which negates positive communication. Cyber trolling is one of these behaviors, and its emergence has been growing in popularity because of its prevalence and the impact it has on the online interactions and experiences of the users (Kross et al., 2021). Most of the time, cyber trolling has been termed as deliberate online actions that are aimed at stirring, annoying, or disrupting the emotions of other persons, usually with the aim of having fun or satisfying oneself (Hardaker, 2010). In contrast to reactive types of online aggression, trolling

is generally intentional and strategic, and the aggressors want to provoke a reaction and retain control over the online communication process. One of the studies indicates that both situational affordances of the online environment and personal personality differences contribute to the development of trolling behaviors (Buckels et al., 2014). Due to the increased integration of online communication into social, educational, and work settings, the psychological peculiarities related to these behaviors have gained particular significance. Personality psychology offers a viable framework that can be used to study the differences in individual behavior in the online environment. Narcissism is one of the personality traits that has been heavily examined regarding digital self-presentation and interpersonal behaviors. Narcissism is a personality that is defined by grandiosity, entitlement, self-concentration as well as a firm need to be admired (Morf & Rhodewalt, 2001). According to the research, people with a more pronounced set of narcissistic characteristics are more active on the Internet and more likely to commit themselves to those actions that help to attract attention and recognition (Buffardi & Campbell, 2008; McCain & Campbell, 2018). Digital spaces, where exposing oneself and self-popularization are frequently rewarded, can thus offer an environment where narcissistic tendencies go more easily vented. Online interaction is very different from face-to-face interaction, especially because it involves less use of nonverbal cues, physical distance, and a distorted social feedback process. Such characteristics affect the way one perceives others, comprehends the social cues and governs their behavior. According to communication scholars, computer-mediated communication exchanges may enhance self-centred behaviour and change interpersonal relationships relative to face-to-face situations (Walther, 1996). Also, online communication can create psychological distance that could decrease emotional connection with other people, and this can affect the user in responding to disagreement, criticism or conflict (Suler, 2004). Self-related psychological processes and empathy are important in the process of directing social behavior in different contexts. Empathy helps people see through others and care about their feelings, whereas self-esteem determines the response to social assessment and responsiveness (Davis, 1983). The processes can be modified by the lack of instant emotional feedback and the asynchronous quality of communication in the digital world. Certain scholars have proposed that a change in digital communication patterns can be related to the empathic engagement change with time (Konrath et al., 2011). The analysis of the intersection of personality traits, emotional processes, and structures of online interactions can be very beneficial to the understanding of general trends in the behavior in the online environment and help to form a more global picture of social behavior in a modern online environment. Besides the general personality variables, scholars have begun to pay closer attention to the problem of online anonymity as a critical situational variable that dictates antisocial online behavior. The internet environment generally allows users to conceal/distort their identities and, hence, are less accountable and socially controlled, and therefore, they do not subjugate social norms that govern how people interact with each other. It has been ascribed to perceived anonymity, which is linked with the online disinhibition effect, where individuals feel less inhibited and more prepared to do malicious or provocative activities online compared to when they are interacting with each other (Suler, 2004). Research has indicated that anonymity can somehow counterbalance personality traits such as narcissism and enhance dominance, sense of entitlements and exploitative personality in the online world (Christopherson, 2007). This makes narcissistic individuals particularly making them more likely to indulge in cyber trolling when anonymity is guaranteed, and the social cost of doing so is low. Another important psychological process that can be used to explain the behavior of cyber trolling is self-esteem. Although there is a general assumption that narcissism is a manifestation of exaggerated self-perception, empirical evidence has revealed that the narcissistic individuals may have weak or conditional self-esteem and therefore, be highly sensitive to the perceived threat, censure or lack of adoration (Kernis & Sun, 1994; Morf & Rhodewalt, 2001). Defensive or offensive responses are easy to go on the offensive during such sensitivity, and

the reaction is visible, immediate and ambiguous in online space. Cyber trolling is therefore a compensatory effect, allowing individuals to subdue one another, or regain self-esteem, or attention through provocation (Buckels et al., 2014). This means that self-esteem may create a mediating effect between narcissistic personality and antagonistic behaviour on the internet. Besides self-related processes, digital empathy is important in controlling online interactions. Digital empathy is the ability to perceive, communicate, and react to the feelings of others in the computer-mediated setting in an appropriate manner (Rueda et al., 2015). Lack of nonverbal communication, including facial expressions and tone of voice, may constrain the accuracy of empathy and decrease the emotional connection with other individuals. Research shows that people with low empathy scores are more apt to experience online harassment and trolling, as less empathy for the emotional situation of others reduces moral and social inhibition (Ang & Goh, 2010).

Narcissism was always linked to a lack of empathic concern, and there is a possibility that this area is where narcissistic traits might play a role in trolling since they limit emotional responsiveness on the internet (Watson et al., 1984). When combined, these results emphasize the need to use an integrative approach to consider both dispositional and situational factors in cyber trolling. Narcissism can leave people prone to attention-seeking and dominance-related behaviour, anonymity, vulnerability to self-esteem, and a lack of digital empathy, which determine the manifestation of these dispositions on the Internet. The interconnected mechanisms are critical to understand to create theoretical frameworks of online misconduct and to inform the interventions targeting the creation of healthier environments in digital communication. Through the analysis of the psychological mechanisms of cyber trolling, a researcher can make an input to the overall initiative of promoting accountability, empathy, and positive interactions within increasingly digital social environments. The dramatic growth of digital means of communication and social media platforms has served to increase cyber trolling, a modern online antisocial behavior, which involves the deliberate act of provocation, harassment, and disruption of internet communication. Cyber trolling has been linked to both negative psychological impacts on victims, such as emotional distress, anxiety, poorer well-being and broader negative effects on online groups such as lowered trust and worse quality of discourse (Buckels et al., 2014; March & Steele, 2020). Despite the increasing academic interest in cyber trolling as a research topic in the fields of psychology and cyber behaviour, the literature in these areas has illuminated little about the psychological mechanisms of trolling behaviour, as most of the literature has concentrated on behavioral descriptions or direct personality correlations. Specifically, narcissism is also a key personality indicator of online antagonism, as it is linked to grandiosity, entitlement, attention seeking, and threatened ego (Paulhus & Williams, 2002). According to empirical evidence, narcissistic individuals tend to commit online aggression and trolling behaviour (Buckels et al., 2014). Nonetheless, much of the existing literature has been based on direct-effect models, which provide little insight into how narcissism is translated into cyber trolling within the unique psychological and structural conditions of the online environment. Digital spaces are inherently different in contrast to in-person interactions, especially with perceived anonymity, lack of social responsibility and a distortion of emotional information. The Online Disinhibition Effect states that anonymity may undermine social control and enable the release of aggressive or deviant responses that individuals may inhibit in the real world (Suler, 2004). Such conditions may be particularly prone in narcissistic people, where anonymity gives a chance to gain attention and demonstrate his/her power without fear of losing reputation to such actions. Besides, narcissism is often associated with vulnerable or contingent self-esteem, as opposed to high stable self-worth, which means that narcissistic persons tend to be more responsive to perceived criticism or withdrawal (Kernis, 2003). Threatened egotism is a theory of aggression that implies aggression is usually a reaction to threats to an inflated yet uncertain self-concept, and cyber trolling may serve as an ego-

protective mechanism on the internet (Baumeister et al., 1996). Also, studies have shown that narcissism is conjoined with a lack of empathy, especially affective empathy, and this can also contribute to the ease of online aggression when deprived of emotional feedback and social influence (Ang et al., 2011; Watson et al., 1984). Although these factors are theoretically relevant, there has been very little analysis of perceived anonymity, self-esteem, and digital empathy as mediating factors explaining the narcissism-cyber trolling relationship. This is not an integrated mediation framework, which is a critical gap in literature. This gap is critical to developing theoretical models of online antisocial behavior and to inform the evidence-based interventions, digital literacy programs, and platform-level policies to decrease cyber trolling and enhance the health of online interactions.

The rising issue of cyber trolling is a big challenge to individuals, online communities, and online platforms, and thus, there is a need to understand the psychological processes underlying such behaviour. Although previous literature had already determined narcissism as a personality association of online aggression that is salient, the available literature has mostly focused on direct connections, which restrict theoretical specificity and practical utility (Buckels et al., 2014; March & Steele, 2020). Considering the unique features of online contexts, such as perceived anonymity and less social interaction, it is necessary to investigate the underlying mechanisms in which narcissistic tendencies would be sublimated into cyber trolling. The research addresses the call for more integrative models of online antisocial behavior to look beyond surface level relationships and provide more explanatory models based on theories of personality and cyber psychology (Morf & Rhodewalt, 2001). Narcissism is a personality construct that is complex and multidimensional in nature, in that it is represented by grandiosity, self-centeredness, entitlement, a significant need to be admired, and a lack of empathy. The concept was derived from the psychoanalytic theory, and it was originally conceptualized as extreme self-love and sick self-centeredness (Freud, 1957). Modern-day psychology. During the timeline concept of narcissism has been redefined by modern-day personality psychology to be a trait, along a continuum, between normative and adaptive forms and maladaptive and pathological forms. Approaches to the present theoretical models are relatively clear in distinguishing between grandiose narcissism and vulnerable narcissism. Dominance, exhibitionism, inflated self-esteem, entitlement, and interpersonal exploitation are some of the characteristics linked to grandiose narcissism. Grandiose narcissists are more likely to be confident, assertive, and socially bold, as well as express antagonism and lack of concern towards others (Back et al., 2013). On the other hand, vulnerable narcissism is defined with emotional instability, extreme sensitivity to criticism, defensiveness, anxiety and unbalanced self-esteem. Despite the seeming introversion or inhibition of the vulnerable narcissists, they have the same underlying entitlement and self-centeredness that were seen in the grandiose narcissism (Pincus & Lukowitsky, 2010). Trait-based approaches also develop the conceptualization of narcissism as several components, which are intertwined and include some superiority beliefs, self-enhancement, entitlement, and empathy impairment. The main feature of narcissism is the process of self-regulation, which is to develop a positive and superior image of oneself. When threatened by this self-image, narcissistic persons tend to use more maladaptive coping mechanisms to recover their self-esteem, such as aggression, hostility or derogation of others (Bushman and Baumeister, 1998). They are mainly ego-defensive reactions that are especially strong in people with weak or conditional self-esteem, which indicates a tight linkage between narcissism and the control of self-esteem. Notably, recent literature has stressed the contextual sensitivity of narcissism, where the behavioral manifestation of narcissism is influenced by the affordances of the environment. Online and in the digital form, low levels of social cues, perceived anonymity, and little accountability may amplify narcissistic behavior because narcissists possess the power to promote themselves, seek out validation, and act aggressively with fewer immediate repercussions (Campbell & Miller, 2011). Narcissism is therefore becoming

a conceptualized phenomenon as both a trait factor as well as a dynamic construct that starts to interact with situational factors and is therefore of particular concern in online behaviors such as cyber trolling and digital aggression. Altogether, narcissism can be best regarded as a complex personality construct which involves intrapersonal and interpersonal dysfunctions. Its modern-day conceptualization incorporates personality traits, self-esteem control mechanisms, lack of empathy, and contextual factors, and has offered a well-developed approach to studying the way in which narcissistic tendencies can be expressed in offline and online social contexts. Cyberbullying is often an expression of recurrent aggression against a particular target and is driven by power disparity or enmity, cyber trolling is often driven by entertainment, attention-seeking, or psychological satisfaction in disrupting the target (Hardaker, 2010). Trolling can be directed at an individual, a group of people or the whole online community, and it is often made easy thanks to the characteristics of the digital platform being anonymous, asynchronous, and less socially accountable. Psychology attributes the causes of cyber trolling to maladaptive personalities and a lack of affect. Empirical studies always correlate trolling with discoveries of the Dark Tetrad, especially, narcissism, Machiavellianism, psychopathy, and day-to-day sadism (Buckels et al., 2014). The central aspect in the conceptualization of cyber trolling is the lack of empathy, callousness, and lack of emotions that allow people to feel no care about the emotional impact of their actions on others. In general, trolls tend to experience disrupted cognitive and affective empathy, which enables them to dehumanize victims and legitimize abusive online behavior (Sest & March, 2017). The contextual issues are very important in determining cyber trolling patterns. Online disinhibition effect indicates that anonymity and invisibility lessen self-control and enhance the probability of demonstrating socially inappropriate behaviors on the internet (Suler, 2004). A sense of anonymity undermines the normative restrictions and moral responsibility and allows individuals to troll without any fear of societal punishment. Furthermore, online space restricts nonverbal communication and emotional feedback and weakens emphatic communication and strengthens aggressive or provocative communication patterns. In general, cyber trolling can be understood as goal-oriented, psychologically oriented cyber-aggression due to the personality and lack of empathy and contextual opportunities of the online environment. The conceptualization of cyber trolling as a disposition (e.g., narcissism), mediated by mental processes (e.g., self-esteem regulation and empathy), and influenced by online conditions (e.g., anonymity), gives a solid theoretical context for analyzing its prevalence and influence in the modern online environment.

Antagonistic and dominance-oriented interpersonal behaviors have always been associated with narcissism, which is a trait marked by grandiosity, entitlement and excessive admiration (Morf & Rhodewalt, 2001). The absence of accountability and more chances to seek attention and provoke are the best circumstances provided by online environments to express such traits. There is empirical evidence pointing out that people who participate in cyber trolling have higher levels of narcissism, meaning that trolling can be used as a form of entertainment, ego reinforcement, and social dominance in the online environment (Buckels et al., 2014). Other recent studies also reinforce this link by showing that narcissism is predictive of online trolling and other disruptive behaviors on social media. Kircaburun et al. (2019) discovered that the participants with high levels of narcissism became more prone to online trolling in situations where they could interact with others impulsively and uncontrollably. On the same note, March and Steele (2020) established that trait narcissism could predispose trolling behavior and that the role of reduced empathy was an important explanatory factor. The same findings imply that narcissistic people might be less concerned with the emotional implications of their online behaviors escalating the risks of intentional provocation and harassment. Simultaneously, the literature points out that the manifestation of narcissistic traits in the online environment is conditioned by the affordances of a situation:

anonymity and a low social presence. In cases where the identifiability and social consequence of online interactions are minimal, the superiority and ego protection motivated narcissism tend to be revealed through antagonistic behaviors (Christopherson, 2007). This process is theoretically described through the Online Disinhibition Effect, which accentuates how the feeling of anonymity and the lack of social presence undermine self-regulation and social norms supporting the aggressive and disruptive online behavior (Suler, 2004). The modern view of digital behavior also indicates that, as the personality traits interact with the social media affordances and generate various responses among the users, the individual vulnerability to the online environment (Valkenburg et al., 2022). In this perspective, narcissists are more sensitive to ego threats and driven to seek attention and dominance, potentially and literally so in cyber trolling in environments that permit disinhibition online. Combined, both the traditional and contemporary empirical findings show a positive relationship between narcissism and cyber trolling behavior, not to mention that the online contexts also increase the behavioral manifestation of narcissistic personality traits. This literature offers a good empirical and theoretical basis to postulate that there is a positive relationship between narcissism and cyber trolling behavior. According to the extant literature on psychology, self-esteem is a key process that mediates the narcissist characteristics in the development of maladaptive interpersonal behaviors such as cyber trolling. Though the concept of narcissism is commonly mistaken with high self-esteem, according to the theories, the narcissistic self-esteem is unstable and contingent, and it is highly reliant on outside validation (Morf & Rhodewalt, 2001). Narcissistic people always need confirmation and, in this way, they are easily influenced by the feeling that they are being criticized, rejected or not given adequate attention. In the event that they fail to get such validation, their self-esteem is at stake, which causes them to develop defensive responses directed at defending or rebuilding their self-worth (Kernis & Sun, 1994). Online spaces exacerbate these self-esteem processes by making people prone to social assessment, comparison in the open, and vague feedback. The likes, the comments, and the reactions are often signs of social approval, and their absence can be considered a threat to self-esteem. Here, narcissistic people might feel more vulnerable as far as their ego is concerned, primarily when they expect to be admired and the expectations remain unmet. The studies on threatened egotism indicate that any person who has unstable self-esteem is more apt to react positively through hostility and violence to ego-threats especially when social restraints are in place and where the social restraints are weak (Baumeister et al., 2003). Cyber trolling can therefore be seen as a detrimental coping mechanism that enables one to reclaim control, dominance or superiority as a response to threatened self-esteem. Empirical evidence gives more credence to the mediating effect of self-esteem in the narcissism-cyber trolling relationship. Bushman and Baumeister (1998) showed that narcissism and ego threat are predictors of aggressive behavior and control by self-esteem facilitates the expression of narcissistic tendencies into hostility. Applying the same frame to online environments, Kircaburun et al. (2019) discovered that narcissism had an indirect correlation with online trolling behavior via self-regulatory weaknesses that were connected to self-esteem. This is because these findings indicate that narcissistic people are not necessarily aggressive but more likely to be antagonistic when their esteem is upset. The most recent studies have strengthened this view by showing that online trolling can serve an ego-defensive purpose for narcissistic people. March and Steele (2020) found that trolling behaviors were predicted by narcissistic traits due to dysfunctional self-regulation, which showed that people could use online provocation to respond to threatened self-worth. Anonymity and lowered accountability on the internet further encourage this process by decreasing the perceived social cost of antagonistic behavior. Cyber trolling is an easy and risk-free option in such situations, which is aimed at regaining the feeling of power and self-importance. Collectively, the literature indicates that self-esteem is one of the major psychological mechanisms that mediate narcissism and cyber trolling behavior. Narcissistic characteristics increase susceptibility to ego threat, whereas

Internet settings enhance self-esteem changes and decrease restraints of behavior. Cyber trolling can also occur as an act of compensation by narcissistic individuals whose self-esteem has been threatened, since they are attempting to safeguard their self-concept. Therefore, self-esteem would not be only a correlational variable, but a working mediating process, which would justify the mediation role of narcissism into disruptive online behavior based on theoretical and empirical foundations, very well.

Empathy has been commonly accepted to be one of the most important psychological processes that shape interpersonal action, especially those related to aggression and antisocial behaviour. It is always true that narcissism is a state that lacks empathy, particularly affective empathy, because narcissists are not concerned with the needs of others, their feelings, or their image among others (Watson et al., 1984; Morf & Rhodewalt, 2001). This weakens the ability to detect or empathize with the feelings of others by restricting moral restraint and augmenting the chances of actions that trigger damage or pain. Within online spaces, where facial expression and the tone of voice are mostly absent, empathic interaction is further eroded, aggravating the impacts of already existing empathy shortcomings. Empathy studies have shown that low empathy is a major predictor of cyber trolling and other similar types of online harassment. The people who are less empathically concerned are not so sensitive to the emotional impact of their behavior and thus are more ready to provoke, offend, or exploit people to satisfy their own needs (Ang & Goh, 2010). Cyber trolling, characterized by intentional efforts to generate emotional responses, can be especially attractive to narcissistic people who do not care about and want attention, dominance, or humor at the expense of others. This pathway is supported by empirical evidence, which shows that narcissism makes predictors of trolling behaviors, and they are predicted by lower levels of empathy (Sest & March, 2017).

The mediating effect of empathy in the relationship between narcissism and trolling is also supported by even more recent studies. March and Steele (2020) discovered that narcissistic characteristics related to trolling behavior (primarily by lesser empathy), which implies that the lack of empathy may be used to explain why narcissistic individuals troll intentionally. On the same note, studies on cyberbullying also emphasize that empathy is a protective element, and its lack eradicates a relevant emotional shield against harmful online conduct. In the disinhibited online space, lack of empathy, combined with anonymity, facilitates the process of dehumanization as well as the lack of emotional connection to the consequences of the behavior towards others (Suler, 2004). Collectively, the readings imply that narcissism does not directly cause cyber trolling in a vacuum and enhances the chances of this behavior through the disrupted functioning of empathy. The anonymity of the Internet enhances identifiability and accountability, undermining social norms and causing a lack of care about negative judgment or penalty. Consequently, they can become freer to communicate hostile and provocative or disruptive behaviors which would have been suppressed in the face-to-face communications (Christopherson, 2007). This relationship can be explained using theoretical approaches to the concept of Online Disinhibition Effect, which is based on the idea that anonymity and lack of social presence decrease self-control and influence impulsivity in online communication (Suler, 2004). Feeling anonymous, people have less self-awareness and moral restraint, and in this case, tend to break social norms. This belief is corroborated by empirical studies, which have shown that antisocial behaviors, such as harassment and trolling, are much more common in the online environment that has the provision of anonymity or pseudonymous communication (Cheng et al., 2017). The support of perceived anonymity in cyber trolling behavior is also supported by recent studies. Studies show that anonymity in the internet environment leads to deindividuation and the encouragement of the user to leave their personal responsibility and empathic concern in a social context, making it more likely to participate in aggressive or provocative behavior

(Sest & March 2017). Furthermore, people who experience the sense of increased anonymity tend to believe that trolling is a low-determination and socially acceptable behavior, which escalates its frequency and severity. These results indicate that anonymity does not simply go hand in hand with cyber trolling but supports and enhances it. Perceived anonymity provides a setting where trolling behavior is most likely to develop and continue because it reduces accountability, undermines social norms, and enables a setting where participants feel less inhibited to interact. This fact shows that anonymity is a key contextual consideration in the explanation of cyber trolling in online space. The perceived anonymity online will be central to the development of the expression of personality traits in the online context, especially as far as antisocial behaviors like cyber trolling are concerned. According to personality theorists, dispositional traits have more power of behavior effects in situations where social norms and accountability become weaker. Internet spaces where a user can hide their identity lessen individual responsibility and lessen the issue of social judgment, which generate the circumstances in which suppressed personality propensities have a greater opportunity to express themselves (Christopherson, 2007). Such contexts can be used by individuals with a high level of narcissistic traits who are characterized with grandiosity, entitlement, and the need to dominate other people to express antagonistic behaviour in the quest to be heard or to feel superior. This moderating process is fully explained by the Online Disinhibition Effect. This framework states that there is less self-awareness and self-monitoring in anonymity and thus less inhibition, more impulsiveness in online communication (Suler, 2004). With perceived high anonymity, narcissists can be more likely to feel freer and less afraid of adverse outcomes, and cyber trolling will be a low-stakes approach to reach ego-related aims. However, in the opposite scenario, where identity transparency or high social regulation is enforced in online environments, narcissistic tendencies can be channeled into socially acceptable types of self-promotion and not blatant provocation. The moderating power of anonymity in the correlation between narcissism and cyber trolling is proven by empirical studies. Research on the relationship between dark personality traits and antisocial behavior online has revealed that the connection between the two exists in anonymous or low-moderation settings (Buckels et al., 2014). In the same way, extensive studies of online communities reveal that antisocial actions, such as trolling, are more likely to happen when the users believe that there are low levels of accountability and low levels of social penalty (Cheng et al., 2017). These results indicate that anonymity only intensifies the behavioral manifestation of the narcissistic predilection, other than providing its independent causal power toward trolling.

The current research is pegged into a theory of Online Disinhibition Effect, which tries to explain the ways and reasons as to why people act in different ways online than how they would act in real life (Suler, 2004). The theory holds that certain structural and psychological peculiarities of digital communication minimize behavioral constraints that traditionally govern the behavior in society, which exposes the probability of unregulated, impulsive, or deviant behavior. These characteristics are perceived anonymity, invisibility, fewer authority cues, synchronicity and less social accountability. All this together changes self-regulation mechanisms and undermines the observance of social norms. In the modern online world, social networking sites, forums, and commenting, people tend to communicate without knowing each other or receiving immediate replies. The Online Disinhibition Effect holds that this low social presence decreases self-awareness and responsibility, which gives people the freedom to think and act in a way that would not normally be exhibited in the physical world. This causes actions like aggression, instigation, and deliberate interference to increase in the online environment (Suler, 2004). Cyber trolling, as a purposeful effort to annoy, aggravate, or emotionally upset people, can hence be interpreted as a direct expression of disinhibition on the internet.

Online Disinhibition Effect gives an effective theoretical explanation of cyber trolling due to the intrinsic connection between trolling behaviors and digital environments by offering anonymity and psychological distance. Studies have always indicated that antisocial online behaviours tend to increase more in environments where users feel that there are few repercussions to their actions (Cheng et al., 2017; Christopherson, 2007). When people believe that they are not under threat of identification and social sanctions, they tend to become more prone to norm-deviating behavior, such as harassment, aggression, and manipulation. Therefore, cyber trolling is not only an act of personal predisposition, but it is also enabled and enhanced by disinhibiting characteristics of the Internet space. In this context, narcissism is theorized as characteristic aspect whose behavioral manifestation depends on the situational limitations. Narcissistic people are distinguished by pomposity, feelings of entitlement and cravings for admiration. Although such characteristics can be socially controlled in the offline setting because of the fear of negative judgment, in online settings, which are characterised by anonymity, such restrictions are minimized. The Online Disinhibition Effect describes how narcissistic personalities become less inhibited in the online environment and more prone to antagonistic actions like cyber trolling. This opinion is supported by empirical data showing that narcissistic personality characteristics correlate with more online aggression and trolling behaviour, especially in an anonymous or minimally restricted environment (Buckels et al., 2014).

The theory also explains how self-esteem causes cyber trolling because it highlights the conditions of disinhibition that place stress on self-regulatory processes. Online communication is usually accompanied by collective appraisal, social comparison, and uncertain feedback, which may increase the ego threat. Depending on the environment, people can behave impulsively when they perceive themselves to be threatened because no immediate social repercussions forcing them to act and thus restrain themselves. Cyber trolling can therefore be a maladaptive behavior to regain a feeling of control, dominance or validation. In terms of the Online Disinhibition Effect, this kind of conduct is an indication of a derailed self-monitoring and an inability to control emotions in the online environment (Suler, 2004). In the same vein, the theory offers a very convincing view on the aspect of empathy in the act of cyber trolling. Anonymity is one of the main elements of the Online Disinhibition Effect and is directly included in the current study as a moderating variable. Anonymity decreases identifiability and personal responsibility thus increasing the effect of personality traits on behavior. The theory, based on the assumption of perceived anonymity, assumes that the more anonymity is perceived, the less internal restraints and the less significant underlying dispositional tendencies can be (narcissism, low empathy, and vulnerability to self-esteem). This is the reason as to why cyber trolling is more prevalent in the anonymous or pseudonymous sites, and the reason why the correlation between narcissism and trolling should be more profound when the condition of high anonymity exists (Christopherson, 2007). The sample population of the current research consists of active internet users who actively use online platforms, which include social networking sites, discussion forums and interactive online communities. This population is suitable since the phenomenon of cyber trolling is present in the online space and can only be studied significantly among the people who are actively involved in online communication. Some of the studies on online aggression and trolling conducted previously targeted online users in general and not a certain professional or demographic setup, as trolling practices are not limited to a particular type of user (Buckels et al., 2014; Cheng et al., 2017). The targeting of active internet users thus improves the ecological soundness of the research, and the respondents are adequately exposed to online communication settings where the trolling behaviors can be practised.

Research Methodology

Under the non-probability sampling, purposive sampling is used as the main sampling method. Purposive sampling will allow the researcher to select the participants who fit the research goals by choosing those who use social media or online discussion platforms frequently (Bryman, 2016). The given methodology is suitable for the current study since it guarantees that the participants were fully acquainted with online interactions to report properly on the matters of cyber trolling behavior, perceived anonymity, and the associated psychological processes. The purposive sampling technique has been adopted widely in the literature related to the study of online misconduct, personality, and digital behavior since it is appropriate when a researcher wants to study information-rich cases (Kircaburun et al., 2019). The study took a sample of 596 respondents that was deemed sufficient and powerful to conduct a quantitative analysis in terms of mediation and moderation tests. Besides the statistical factors, the adopted sample size enhances the external validity of the study due to the inclusion of variability of personality traits, psychological mechanisms, and online behaviors among a wide variety of users. The current research adopted a survey research design using a designed questionnaire that was conducted on an electronic Google Form to gather information from the respondents. The data was collected electronically through Google Forms that enabled it to reach a broad geographic area, distribute data cheaply and efficiently manage it. Narcissism, the independent variable, is operationalized as a subclinical personality trait reflecting grandiosity, entitlement, self-importance, and a desire for admiration. It is measured using nine items, adapted from the Narcissistic Personality Inventory short form (NPI-16). Items such as “People see me as a natural leader,” “Many group activities tend to be dull without me,” and “I insist on getting the respect I deserve” capture core narcissistic tendencies. Responses are recorded on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Negatively worded items are reverse-coded before analysis, and higher composite scores indicate higher levels of narcissism (Ames et al., 2006).

Cyber trolling behavior, the dependent variable, is operationalized as intentional online behavior aimed at provoking, irritating, or emotionally disturbing others. This construct is measured using twelve items assessing self-identification with trolling, enjoyment of trolling, alignment with trolling culture, and attitudes toward trolling behavior. Sample items include “I consider myself to be a troll,” “I enjoy trolling other people,” and “I identify with trolling culture.” Items reflecting disapproval of trolling are reverse-coded. Responses are measured on a five-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree), with higher scores indicating greater engagement in cyber trolling behavior (Buckels et al., 2014).

Self-esteem, the first mediating variable, is operationalized as an individual’s overall evaluation of self-worth and self-acceptance. It is measured using ten items adapted from the Rosenberg Self-Esteem Scale. Items include “On the whole, I am satisfied with myself,” “I feel that I am a person of worth,” and “I take a positive attitude toward myself,” along with negatively worded statements such as “I feel useless at times.” Responses are recorded on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Negatively phrased items are reverse-coded, and higher scores reflect higher levels of self-esteem (Rosenberg, 1965).

Empathy, the second mediating variable, is operationalized as the capacity to understand and respond to others’ emotional experiences. Empathy is measured using sixteen items adapted from the Interpersonal Reactivity Index, focusing primarily on emotional concern and affective responsiveness. Sample items include “I enjoy making other people feel better,” “I can tell when others are sad even when they do not say anything,” and “I feel kind of protective when I see someone being taken advantage of.” Responses are recorded on a four-

point Likert scale ranging from 1 (Never) to 4 (Often). Negatively worded items are reverse-coded, and higher scores indicate greater empathic capacity (Davis, 1983).

Perceived online anonymity, the moderating variable, is operationalized as individuals' subjective perception of being unidentifiable and free from personal accountability in online environments. It is measured using six items, including statements such as "It is difficult for others to identify me as an individual" and "I believe that my personal identity remains unknown to others." Items reflecting identifiability are reverse-coded. Responses are measured on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), with higher scores indicating greater perceived anonymity (Christopherson, 2007).

Data Analysis

Table 4. 1 : Demographics

Variable	Frequency (n)	Percentage (%)
Gender		
Male	266	44.6
Female	321	53.9
Other	9	1.5
Age		
Mean ± SD	26.62 ± 8.82	—
Education		
Intermediate	60	10.1
Bachelor's	147	24.7
Master's	231	38.8
MPhil	130	21.8
PhD	28	4.7
Marital Status		
Single	333	55.9
Married (including widowed)	218	36.6
Divorced	45	7.6
Location		
Urban	596	100.0
Monthly Income (PKR)		
Below 50,000	196	32.9
50,001–100,000	173	29.0
100,001–150,000	165	27.7
150,001–200,000	40	6.7
Above 200,000	22	3.7

The sample size was 596, and the gender representation was slightly disproportionate (53.9 and 44.6 percent respectively) with a minor percentage of other (1.5 percent). The age of the participants was 26.62 (SD = 8.82), which shows that the sample was composed mostly of young adults, whose ages were between 18 and 60 years. Regarding education, most of the participants were well educated, with 38.8% having a master's degree and 21.8% having an MPhil qualification. This indicates that the sample is academically informed and, specifically, it is applicable in comprehending the online behavioral pattern of digitally engaged people. Concerning marital status, more than half of the respondents were single (55.9), and 36.6 were married and widowed, and 7.6 were divorced. This distribution is representative of a mostly unmarried sample, which is consistent with the younger age composition of the sample. This means that all the participants (100 percent) were based in urban locations; therefore, the results can best represent urban communities and may not be applicable in rural locations. About the income, most of the respondents are low to middle income earners and those who are below PKR 50,000 earners are 32.9 percent, those who are on PKR 50,001 to 100,000 earners are 29.0 percent and those on PKR 100,001 to 150,000 earners are 27.7 percent. Only a few percent of the participants indicated increased income levels.

Table 4. 2 : Internal Consistency

Variable	No. of Items	Cronbach's α
Narcissism	9	0.7
Cyber Trolling	12	.73
Anonymity	6	.63
Self-Esteem	10	.76
Digital Empathy	16	.78

All scales of the study were estimated using alpha coefficients of Cronbach to evaluate their internal consistency. Cyber Trolling scale had acceptable reliability ($\alpha = .73$), whereas Self-Esteem ($\alpha = .76$) and Digital Empathy ($\alpha = .78$) were good in terms of internal consistency. The scale on narcissism produced a reliability coefficient of $\alpha = .70$, which is acceptable in psychological studies, especially when the construct being studied is a personality that expresses itself in different ways. On the same note, the scale of Anonymity had a Cronbach's alpha of $\alpha = .63$, and this is within the acceptable range of research, exploratory and social sciences, particularly when the number of items is comparatively low. In general, reliability coefficients demonstrate that all the instruments applied to the research have a good to adequate level of internal consistency, which can be later used in further analyses, such as correlation and mediation analysis.

Table 4. 3: Descriptive Statistics

Variable	Mean	SD	Min	Max
Narcissism	3.174	0.745	1.00	5.00
Cyber Trolling	3.101	0.678	1.00	5.00
Self-Esteem	3.580	0.765	1.00	5.00
Empathy	2.943	0.597	1.00	5.00
Anonymity	3.306	0.811	1.00	5.00

The descriptive statistics reveal that the respondents had an intermediate level of study variables. The mean score of narcissism was at 3.17 (SD = 0.75), indicating that the subjects, on average, had moderate to slight tendencies of narcissism. Another moderated mean was

that of cyber trolling (M = 3.10, SD = 0.68), which meant that participation in or support of trolling actions was in the middle of the sample. Self-esteem has shown the greatest mean score of the variables (M = 3.58, SD = 0.77), which means that the respondents tended to have a positive assessment of themselves and personal value. Empathy showed a relatively lower mean (M = 2.94, SD = 0.60), which means that the participants showed moderate empathic behaviours. Online perceived anonymity exhibited a mean of 3.31(SD=0.81), which depicts a moderate perception of anonymity of identity in online space. All in all, the standard deviations show sufficient differences in all the constructs, which indicates significant individual differences in the sample. The fact that the observed values do not have extreme skewness (as evidenced by the whole range of the obtained scores 1-5) indicates that the data is adequate and can be further used in the inferential tests, i.e., correlation, regression, and mediation.

Pearson correlation analysis revealed significant associations among all study variables ($p < .001$). Narcissism was positively correlated with cyber trolling ($r = .574$), self-esteem ($r = .496$), and perceived anonymity ($r = .488$), while showing a negative relationship with empathy ($r = -.413$). Cyber trolling was positively associated with self-esteem ($r = .538$) and anonymity ($r = .567$), but negatively related to empathy ($r = -.459$). Additionally, self-esteem demonstrated a positive correlation with anonymity ($r = .579$) and a negative correlation with empathy ($r = -.523$). Anonymity was negatively associated with empathy ($r = -.396$). Overall, the pattern suggests that higher narcissism, greater perceived anonymity, and higher self-esteem are linked with increased cyber trolling, whereas empathy appears to function as a protective factor, being inversely related to both narcissism and trolling behavior.

Table 4. 4: Correlation Matrix

Variable	1	2	3	4
1. Narcissism	1.000			
2. Cyber Trolling	0.574***	1.000		
3. Self-Esteem	0.496***	0.538***	1.000	
4. Empathy	-0.413***	-0.459***	-0.523***	1.000
5. Anonymity	0.488***	0.567***	0.579***	-0.396***

*** $p < .001$; * $p < .01$; $p < .05$ (two-tailed).

Table 4.5: Simple Regression Analyses Predicting Cyber Trolling

Predictor	B	SE	t	p	R ²
Narcissism	0.574	0.034	16.88	< .001	.329
Anonymity	0.567	0.035	16.20	< .001	.321
Empathy	-0.459	0.034	-13.50	< .001	.211
Self-Esteem	0.538	0.036	14.94	< .001	.289

An examination of the predictive influence of narcissism, anonymity, empathy, and self-esteem on cyber trolling was performed by a series of simple regression analyses. The findings showed that cyber trolling was affected significantly by all four predictors ($p < .001$). The best predictor was narcissism (B = 0.574, R² = .329), which predicted 32.9% of cyber trolling. The positive effect of anonymity was also quite significant (B = 0.567, R² = .321) and explained 32.1 percent of the variance. There was also a positive, but significant influence of self-esteem on cyber trolling (B = 0.538, R² = .289), which accounted 28.9% of the variation. Conversely, cyber trolling had a negative correlation with empathy instead (B = -0.459, R² = .211), which explains the reason why higher empathy levels relate to less trolling

behavior. Altogether, the strongest predictive power was observed with narcissism and anonymity, and empathy was used as a protective factor against cyber trolling.

Table 4. 6: Mediation Analysis: Self-Esteem as a Mediator Between Narcissism and Cyber Trolling (PROCESS Model 4)

Path	Relationship	B	SE	t	p
a	Narcissism → Self-Esteem	0.496	0.036	13.78	< .001
b	Self-Esteem → Cyber Trolling	0.538	0.036	14.94	< .001
c	Narcissism → Cyber Trolling (Total Effect)	0.574	0.034	16.88	< .001
c'	Narcissism → Cyber Trolling (Direct Effect)	0.412	0.051	8.08	< .001

Indirect Path	Effect	Boot SE	95% CI (LL, UL)
Narcissism → Self-Esteem → Cyber Trolling	0.267	0.041	[0.189, 0.352]

The mediating hypothesis to test the relationship between self-esteem and the association between narcissism (independent variable) and cyber trolling behavior (dependent variable) was performed as a mediation analysis with Hayes PROCESS macro (Model 4). The specified predictor was narcissism, the specified mediator was self-esteem, and the outcome variable was cyber trolling. The findings meant that narcissism was a significant predictor of self-esteem (a path), $b = 0.50$, $p < .001$ and self-esteem was a significant predictor of cyber trolling behavior (b path), $b = 0.54$, $p < .001$. The overall impact of cyber trolling due to narcissism was highly noticeable (c path), $b = 0.57$, $p < .001$. Inclusion of self-esteem into the model produced a significant but lower value of the direct effect of narcissism on cyber trolling (c' path), which suggests partial mediation. Notably, the self-esteem mediated narcissism to cyber trolling was statistically significant, since the bootstrapped confidence interval was not zero.

Table 4. 7: Moderated mediation analysis

Path	Relationship	B	SE	t	p
a	Narcissism → Self-Esteem	0.496	0.036	13.78	< .001
a ₃	Narcissism × Anonymity → Self-Esteem	0.021	0.028	0.75	.453
b	Self-Esteem → Cyber Trolling	0.538	0.036	14.94	< .001
c'	Narcissism → Cyber Trolling (Direct)	0.412	0.051	8.08	< .001

Level of Anonymity	Indirect Effect	Boot SE	95% CI (LL, UL)
Low (-1 SD)	0.259	0.042	[0.181, 0.343]
Mean	0.267	0.041	[0.189, 0.352]
High (+1 SD)	0.274	0.045	[0.192, 0.361]

The moderated mediation analysis was performed with the help of Hayes, PROCESS macro (Model 7) to investigate the hypothesis that the perceived online anonymity condition mediates the indirect impact of narcissism on cyber trolling behavior via self-esteem. The findings revealed that narcissism was a very important predictor of self-esteem and vice versa. Narcissism and perceived online anonymity did not have significant interaction effects on the prediction of self-esteem, and this shows that anonymity does not mediate the relationship between narcissism and self-esteem. In addition, the indirect influence of narcissism on cyber trolling via self-esteem was also substantial at low, mean, and high levels

of perceived anonymity online, with the bootstrapped confidence intervals not taking the value of zero. The index of moderated mediation was, however, not significant because the indirect effects were not substantially different at various levels of anonymity. These results indicate that self-esteem mediates the correlation between narcissism and cyber trolling behavior, yet the indirect action is not predetermined by perceived anonymity online. The perceived online anonymity was analyzed by a moderated mediation analysis, which involves Hayes PROCESS macro (Model 7), to determine whether narcissism and cyber trolling behavior have an indirect relationship mediated by empathy. The findings established that narcissism was a major predictor of empathy and empathy was a major predictor of cyber trolling. The interaction between narcissism and perceived online anonymity did not however significantly predict empathy and as such, anonymity does not moderate the relationship between narcissism and empathy. Moreover, the indirect relationship between narcissism on cyber trolling via empathy was statistically significant at the phases of low, mean, and high perceived levels of online anonymity because the bootstrapped confidence intervals were not zero. The index of moderated mediation was not significant because the conditional indirect effects were similar at all levels of anonymity. The results suggest that empathy mediates the association between narcissism and cyber trolling behavior, and such mediation does not depend upon the perceived online anonymity.

Table 4. 8: Moderated mediation analysis

Path	Relationship	B	SE	t	p
a	Narcissism → Empathy	-0.413	0.034	-12.15	< .001
a3	Narcissism × Anonymity → Empathy	-0.018	0.029	-0.62	.536
b	Empathy → Cyber Trolling	-0.459	0.034	-13.50	< .001
c'	Narcissism → Cyber Trolling (Direct)	0.412	0.051	8.08	< .001

Level of Anonymity	Indirect Effect	Boot SE	95% CI (LL, UL)
Low (-1 SD)	0.190	0.039	[0.121, 0.268]
Mean	0.189	0.038	[0.118, 0.261]
High (+1 SD)	0.186	0.041	[0.113, 0.269]

Discussion

The current research aimed to investigate the correlation between narcissism and cyber trolling behavior, through a combination of the personality traits, psychological mechanisms, and the contextual factors, online behavior, into a detailed analytical picture. Particularly, the research examined narcissism as a direct predictor of cyber trolling behavior, mediation of self-esteem and empathy, and moderating and moderated mediation of perceived online anonymity. Using the correlation analysis, the regression analysis, the mediating role and the mediating moderated mediation models when Hayes uses the PROCESS macro, the study can give a finer insight into how the narcissistic tendencies are translated into the antagonistic online behavior. The results obtained show that a narcissistic personality type is a strong positive predictor of cyber trolling behavior. This finding will be correlated with an expanding range of recent studies that confirm that narcissistic personality is closely linked to agile, attention-seeking, and disruptive online behaviors (Kircaburun et al., 2020; March & Marrington, 2021; Sun & Wu, 2023). Narcissistic people are also known to be grandiose and entitled, as well as show a greater desire to be admired, which can drive them to act in ways that elicit emotional responses or establish dominance in virtual space. Social media and forums are the most favored platforms to express narcissism in the form of trolling behaviors because they are immediate feedback platforms, visible, and less socially constrained. In

addition to the direct correlation, the study also established that self-esteem is a mediating variable regarding the association between narcissism and cyber trolling behavior. Self-esteem was positively correlated with narcissism, and self-esteem, in its turn, predicted cyber trolling significantly. This observation corresponds to the modern models, which suppose the conceptualization of narcissistic self-esteem as an unstable and contingent, but not necessarily secure, one (Brummelman et al., 2020; Geukes et al., 2021). Narcissistic individuals might have exaggerated perceptions about themselves, which they need to constantly reinforce, and internet trolling might serve as a compensation mechanism to preserve or protect this vulnerable self-perception. Aggressive or provocative online conduct may become a mode of revitalising a feeling of perceived superiority and dominance over others when self-esteem is conditional on external validation. Empathy was also a considerable worthy mediator between narcissism and cyber trolling behavior. Increased narcissism was linked with decreased empathy that subsequently predicted increased involvement in cyber trolling. The given finding can be attributed to the recent empirical findings indicating that empathy deficits are the key factor in online aggression and digital deviance (Chen et al., 2021; Pabian et al., 2023). Narcissistic people are not very sensitive to emotional damage done to others by their online activities, so they have less emotional concern towards others. Low empathy can also contribute to trolling behavior in the digital setting, which is characterized by constrained social cues and reduced prominence of emotional impact. The mediating value of empathy highlights the relevance of emotions and interpersonal conditions to the explanation of the causes of the development of narcissistic characteristics in the form of online antagonism. In contrast with the conventional assumptions, the perceived online anonymity did not moderate the relationships between narcissism and cyber trolling behavior, nor did it moderate the indirect relationships via self-esteem or empathy. Whereas the perceived online anonymity was found to be a predictor of cyber trolling actions on its own, the strength of the psychological connections between narcissism and trolling did not change. This result contradicts the classical views of deindividuation that propose that the absence of accountability and self-awareness increases antisocial behavior due to anonymity (Zimbardo, 1969). Rather, the findings are consistent with more recent literature that suggests that the role of personality traits is typically more important than situational anonymity in forecasting online delinquency (Kowert et al., 2020; Chen & Chang, 2022).

There are no moderation effects, which implies that narcissistic personalities tend to commit cyber trolling whether they feel that they are anonymous or not. It can be a sign of the further normalization of unfriendly online communication and the absence of stigma in trolling activities on most Internet resources. Furthermore, the fact that the interaction plots show parallel slopes speaks in favor of the conclusion that anonymity does not precondition the effect of narcissism on self-esteem or empathy. Such results indicate that anonymity is an empowering background condition, but not a trigger that changes psychological processes. The moderated mediation analyses also indicated that even though the effects narcissism had on cyber trolling through self-esteem and empathy were significant, they did not change with low, medium, and high perceived online anonymity. It means that the psychological mechanisms that interconnect narcissism and trolling behavior are strong and fairly consistent across various contexts on the Internet. Recent reports have also indicated that the lack of empathy and inappropriate maladaptive self-regulatory processes are predictors of online aggression irrespective of the anonymous levels (Wachs et al., 2022; Schade et al., 2024). Therefore, anonymity reduction can not be a panacea for dealing with cyber trolling unless there is a psychological predisposition that is being dealt with. Taken altogether, the results indicate that the tendency towards cyber trolling is entrenched deeper in the psychological peculiarities of individuals than the anonymity of the situation itself. Narcissism has a positive and negative influence on trolling behavior because it has a direct and indirect

impact on it by means of self-esteem and empathy, which is why we should refer to internal self-regulatory and emotional processes. Although online anonymity is perceived to have a positive effect on cyber trolling, it does not in itself alter the nature of the transposition of narcissistic characteristics into destructive online action. Theoretically, this research will be a continuation of modern models of deviant behavior online since it will incorporate the mediation and moderated mediation models in personality psychology. The results affirm the current demands to go beyond the simplistic views of online aggression that only emphasize anonymity or platform structure (Kircaburun et al., 2020; Kowert & Oldmeadow, 2023). Rather, the findings point to the necessity to factor in stable personality characteristics and psychological processes in explaining cyber trolling. In practical terms, the results have significant implications for intervention and prevention strategies. Other programs to mitigate cyber trolling can be helped by addressing empathy, emotional control, and adaptive self-esteem instead of solely the anonymity reduction or punitive reactions. Educational programs, digital literacy trainings, and psychological assistance programs that can lead to sympathy and the development of a positive self-concept can prove especially useful in alleviating the trolling tendency, particularly in people with a high level of narcissism.

Conclusion

This study investigated the connection between narcissism and cyber trolling behavior with special focus on the psychological processes and situational conditions that could be behind this connection. The study was carried out with a quantitative research design to understand the direct impact of narcissism on cyber trolling, the mediating and moderating influence of self-esteem and empathy, as well as the perceived online anonymity. The analysis of data was based on descriptive statistics, reliability analysis, correlation analysis, regression analysis, and Hayes mediation and moderated mediation model testing, relying on the use of the PROCESS macro. The demographic result of the respondents was a mixed sample, which offered enough variability of the respondents in the critical attributes like age, gender, level of education and internet usage. Such diversity contributed to the strength of the results and made sure that the dataset could be subjected to the inferential statistical analyses. The lack of severe demographic imbalance also contributed to the ability to generalize the findings to a field with a similar population. The correlation analysis showed that there were significant correlations among all major study variables. There was a positive relationship between cyber trolling and narcissism, which implies that an individual with high narcissistic behavior was more likely to practice trolling-related behavior. There were also significant relationships between narcissism and self-esteem, empathy, and perceived online anonymity, indicating that the deviation of narcissism tendencies is associated with the inner psychological mechanisms as well as the perception of the online space. There was a strong correlation between cyber trolling and self-esteem, empathy, and anonymity with empathy coming out as a powerful correlate, indicating its applicability as a protective psychological factor. These correlational results were initial indications of the suggested conceptual framework and warranted additional regression and PROCESS analysis. Simple regression analyses affirmed that the criteria of narcissism, self-esteem, empathy, and perceived online anonymity had individual predictions of cyber trolling behavior. The narcissism and perceived online anonymity proved to be powerful positive predictors, which explains that both personality factors and situational perception drive trolling behavior. Empathy proved to have a substantial negative impact, implying that more empathic interviewees are less prone to cyber trolling. There was also a strong relationship between self-esteem and cyber trolling, which supports the importance of self-evaluative processes that relate to online behavior. Combined, these results point to the fact that cyber trolling depends both on specific personal psychological peculiarities as well as on the contextual factors online. Mediations furnished more information as to the mechanisms behind the relationship between narcissism and the behavior of cyber trolling. The findings showed that there is a partial mediating effect of self-

esteem, i.e. narcissism has both direct and indirect effects on cyber trolling via self-esteem. The association between narcissistic traits and high self-esteem, which in turn predicted cyber trolling behavior, did not exclude the direct impact of narcissism. This partial mediation implies that self-esteem is a key psychological mechanism that connects narcissism to negative behaviour on the Internet. On the same note, empathy mediated the association between narcissism and cyber trolling. An elevated narcissism scores were linked to a lower level of empathy, and a weak level of empathy was a strong predictor of more cyber trolling. These results highlight the key importance of inner psychological mechanisms in changing the characteristics of narcissism into damaging online development. The research also investigated how perceived online anonymity moderates and/or mediates such relationships using moderation and moderated mediation analysis. Even though online perceived anonymity was an independent variable in predicting cyber trolling, it was not a significant modulator of the direct correlation between narcissism and cyber trolling. Also, the moderated mediation analyses showed that perceived online anonymity did not moderate the indirect effects of the narcissism to cyber trolling by either self-esteem or empathy. The significance of its indirect effects using both mediators was significant at all levels of the anonymity, but the magnitude of the strength did not differ significantly at low, medium and high levels of perceived anonymity. These findings were further corroborated by the interaction plots which showed that the slopes were mostly parallel across the levels of anonymity. These findings show that even though anonymity is a significant contextual predictor of cyber trolling, it does not mediate the psychological processes that mediate the relationship between narcissism and trolling behavior.

Limitations & Recommendations

Despite the work done in the present study, there are several limitations that must be mentioned. First, the research design adopted was a cross-sectional design, which limits making a causal conclusion between narcissism, self-esteem, empathy, perceived online anonymity, and cyber trolling behavior. Even though mediation and moderated mediation analyses help to understand the possible mechanisms, the time sequence of variables cannot be determined unambiguously. Future research in this case would be better suited to longitudinal or experimental designs to help understand the causal dynamics and the changes in the psychological processes through time (Kline, 2016; Maxwell & Cole, 2007). Lastly, even though self-esteem and empathy were considered as mediators, other psychological processes like moral disengagement, impulsivity, emotion regulation and social dominance orientation were not modeled. Practically, the implications of the findings propose that interventions that reduce cyber trolling cannot be limited to structural solutions, including anonymity reduction. More effective in addressing the psychological causes of trolling behavior may be programs aimed at improving the level of empathy, emotional regulation, and adaptive self-esteem. Empathy training and self-awareness classes could be added to the educational programs of digital literacy to foster healthier online communication, especially in people with high levels of narcissism (Pabian et al., 2023).

References

- Ang, R. P., & Goh, D. H. (2010). Cyberbullying among adolescents: The role of affective and cognitive empathy, and gender. *Child Psychiatry & Human Development, 41*(4), 387–397. <https://doi.org/10.1007/s10578-010-0176-3>
- Ang, R. P., Tan, K. A., & Mansor, A. T. (2011). Normative beliefs about aggression as a mediator of narcissistic exploitativeness and cyberbullying. *Journal of Interpersonal Violence, 26*(13), 2619–2634. <https://doi.org/10.1177/0886260510388286>
- Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides

- of narcissism. *Journal of Personality and Social Psychology*, 105(6), 1013–1037. <https://doi.org/10.1037/a0034431>
- Baumeister, R. F., Smart, L., & Boden, J. M. (2003). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review*, 103(1), 5–33. <https://doi.org/10.1037/0033-295X.103.1.5>
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Brummelman, E., Thomaes, S., & Sedikides, C. (2020). Separating narcissism from self-esteem. *Current Directions in Psychological Science*, 29(4), 341–347. <https://doi.org/10.1177/0963721420917729>
- Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.
- Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, 97–102. <https://doi.org/10.1016/j.paid.2014.01.016>
- Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression. *Journal of Personality and Social Psychology*, 75(1), 219–229. <https://doi.org/10.1037/0022-3514.75.1.219>
- Chen, L., Ho, S. S., & Lwin, M. O. (2021). A meta-analysis of factors predicting cyberbullying perpetration. *Aggression and Violent Behavior*, 58, 101613. <https://doi.org/10.1016/j.avb.2021.101613>
- Chen, Y., & Chang, C. (2022). Personality traits and online deviant behavior: The role of empathy and moral disengagement. *Computers in Human Behavior*, 129, 107132. <https://doi.org/10.1016/j.chb.2021.107132>
- Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Antisocial behavior in online discussion communities. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 61–70.
- Christopherson, K. M. (2007). The positive and negative implications of anonymity in Internet social interactions. *Computers in Human Behavior*, 23(6), 3038–3056. <https://doi.org/10.1016/j.chb.2006.09.001>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.
- Donnellan, M. B., Trzesniewski, K. H., & Robins, R. W. (2015). Measures of self-esteem. *Handbook of personality assessment*.
- Fox, J., & Moreland, J. J. (2015). The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use. *Computers in Human Behavior*, 45, 168–176. <https://doi.org/10.1016/j.chb.2014.11.083>
- Geukes, K., Nestler, S., Hutteman, R., Dufner, M., Küfner, A. C. P., & Back, M. D. (2021). Puffed-up but fragile: Narcissism, self-esteem, and self-concept clarity. *Journal of Personality and Social Psychology*, 121(5), 1099–1121. <https://doi.org/10.1037/pspp0000365>
- Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, 6(2), 215–242. <https://doi.org/10.1515/jplr.2010.011>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.). Guilford Press.
- Kernis, M. H. (2003). Toward a conceptualization of optimal self-esteem. *Psychological Inquiry*, 14(1), 1–26. https://doi.org/10.1207/S15327965PLI1401_01
- Kernis, M. H., & Sun, C. R. (1994). Narcissism and reactions to interpersonal feedback. *Journal of Research in Personality*, 28(1), 4–13. <https://doi.org/10.1006/jrpe.1994.1002>
- Kircaburun, K., Demetrovics, Z., & Tosuntaş, Ş. B. (2019). Analyzing the links between problematic social media use, narcissism, and online trolling behaviors. *International*

- Journal of Mental Health and Addiction*, 17(4), 891–908.
<https://doi.org/10.1007/s11469-018-9949-8>
- Kircaburun, K., Jonason, P. K., & Griffiths, M. D. (2020). The dark triad traits and problematic online behaviors. *Journal of Behavioral Addictions*, 9(4), 1–14.
<https://doi.org/10.1556/2006.2020.00049>
- Kowert, R., & Oldmeadow, J. A. (2023). Social identity and online aggression. *Current Opinion in Psychology*, 49, 101512. <https://doi.org/10.1016/j.copsyc.2022.101512>
- Kowert, R., Breuer, J., Quandt, T., & Festl, R. (2020). Dark participation in games and online communities. *Computers in Human Behavior*, 105, 106209.
<https://doi.org/10.1016/j.chb.2019.106209>
- Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., & Scherlis, W. (2002). Internet paradox revisited. *Journal of Social Issues*, 58(1), 49–74.
<https://doi.org/10.1111/1540-4560.00248>
- Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., Shablack, H., Jonides, J., & Ybarra, O. (2021). Social media and well-being: Pitfalls, progress, and next steps. *Trends in Cognitive Sciences*, 25(1), 55–66.
- March, E., & Marrington, J. (2021). Antagonistic personality traits and online trolling. *Personality and Individual Differences*, 173, 110613.
<https://doi.org/10.1016/j.paid.2021.110613>
- March, E., & Steele, G. (2020). High esteem and hurting others online: Trait narcissism predicts cyber-aggression through perceived social power. *Aggressive Behavior*, 46(6), 535–545. <https://doi.org/10.1002/ab.21910>
- McCain, J. L., & Campbell, W. K. (2018). Narcissism and social media use: A meta-analytic review. *Psychology of Popular Media Culture*, 7(3), 308–327.
<https://doi.org/10.1037/ppm0000137>
- Morf, C. C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry*, 12(4), 177–196.
https://doi.org/10.1207/S15327965PLI1204_1
- Pabian, S., De Backer, C. J. S., & Vandebosch, H. (2023). Empathy and online aggression: A systematic review. *Aggression and Violent Behavior*, 68, 101809.
<https://doi.org/10.1016/j.avb.2022.101809>
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)
- Rueda, R., Chee, Y. S., & Leong, S. (2015). Developing digital empathy. *Educational Technology Research and Development*.
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (8th ed.). Pearson Education.
- Schade, S., Brugman, D., & Sijtsema, J. J. (2024). Moral disengagement and online antisocial behavior. *Computers in Human Behavior*, 148, 107894.
<https://doi.org/10.1016/j.chb.2023.107894>
- Sest, N., & March, E. (2017). Constructing the cyber-troll: Psychopathy, sadism, and empathy. *Personality and Individual Differences*, 119, 69–72.
<https://doi.org/10.1016/j.paid.2017.06.038>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321–326.
<https://doi.org/10.1089/1094931041291295>
- Sun, Y., & Wu, Y. (2023). Narcissism and toxic online behavior: A meta-analytic review. *Cyberpsychology, Behavior, and Social Networking*, 26(6), 387–396.
<https://doi.org/10.1089/cyber.2022.0413>
- Valkenburg, P. M., Beyens, I., Pouwels, J. L., van Driel, I. I., & Keijsers, L. (2022). Social media use and adolescents' mental health: Toward an individualized differential susceptibility model. *Current Directions in Psychological Science*, 31(3), 243–251.

- Wachs, S., Wright, M. F., & Gámez-Guadix, M. (2022). Social–emotional factors in cyber aggression. *Journal of Adolescence*, *94*, 85–97. <https://doi.org/10.1016/j.adolescence.2021.12.002>
- Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. In M. L. Knapp & J. A. Daly (Eds.), *The SAGE handbook of interpersonal communication* (4th ed., pp. 443–479). Sage Publications.
- Watson, P. J., Grisham, S. O., Trotter, M. V., & Biderman, M. D. (1984). Narcissism and empathy: Validity evidence for the Narcissistic Personality Inventory. *Journal of Personality Assessment*, *48*(3), 301–305. https://doi.org/10.1207/s15327752jpa4803_12
- Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation. In W. J. Arnold & D. Levine (Eds.), *Nebraska Symposium on Motivation* (pp. 237–307). University of Nebraska Press.