

Wang, C.-Y., Liu, Y.-L., & Chang, C.-Y. (2026). The dark triad and cyber aggression: Testing the longitudinal mediation of moral disengagement and toxic online disinhibition. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 20(2), Article 6. <https://doi.org/10.5817/CP2026-2-6>

The Dark Triad and Cyber Aggression: Testing the Longitudinal Mediation of Moral Disengagement and Toxic Online Disinhibition

Cheng-Yen Wang^{1,2}, Yih-Lan Liu², & Chia-Yun Chang²

¹Graduate Institute of Education, Tunghai University, Taichung, Taiwan

²Institute of Education, National Yang Ming Chiao Tung University, Hsinchu City, Taiwan

Abstract

An emerging body of research has consistently linked the Dark Triad traits—Machiavellianism, narcissism, and psychopathy—to cyber aggression. Moral disengagement and toxic online disinhibition have been identified as two psychological processes that may explain this association. However, longitudinal studies simultaneously examining these two mediators in the context of cyber aggression remain scarce, even though such designs are essential for clarifying temporal order and capturing dynamic processes. To address this gap, we conducted a three-wave longitudinal study in which 625 participants completed all three waves of the survey ($M_{age} = 27.54$, $SD_{age} = 6.23$, age range: 20–58; 392 females). We tested a longitudinal mediation model to examine whether moral disengagement and toxic online disinhibition mediate the relationship between the Dark Triad traits and cyber aggression. Toxic online disinhibition was found to be the only significant mediator. This mediating effect was statistically significant for Machiavellianism and psychopathy, but not for narcissism. These findings suggest that toxic online disinhibition may serve as a more robust pathway than moral disengagement in explaining how specific Dark Triad traits contribute to cyber aggression.

Keywords: dark triad; toxic online disinhibition; cyber aggression; moral disengagement; social media user

Editorial Record

First submission received:
November 29, 2024

Revisions received:
March 31, 2025
May 26, 2025
September 4, 2025
October 23, 2025

Accepted for publication:
January 21, 2026

Editor in charge:
Joris Van Ouytsel

Introduction

Cyber aggression is an umbrella term that encompasses various forms of aggressive behaviors intentionally enacted through electronic means to harm individuals (Eun Jahng, 2024). It broadly refers to hostile actions perpetrated via information and communication technologies, including social media platforms and instant messaging applications (C.-Y. Wang & Bi, 2025). These behaviors may occur in direct forms, such as sexual harassment or personal insults, or in indirect forms, such as social exclusion or the dissemination of gossip (Eun Jahng, 2024; Vale et al., 2018). Victims of cyber aggression often experience severe emotional and psychological consequences, including heightened anxiety and depression (Barlett et al., 2024; Eun Jahng, 2024; Kowalski et al., 2014; Sun et al., 2024). Conceptually, cyber aggression differs from other commonly used terms for online aggressive behaviors (e.g., cyberbullying) in that it does not necessarily involve repetition or a power imbalance, rendering its definition clearer and less controversial (Bauman et al., 2013).

Given the substantial impact of cyber aggression, clarifying the role of personality traits is essential for informing the development of effective prevention and intervention strategies (Kowalski et al., 2014). In this regard, the Dark Triad has been widely recognized as a key predictor of maladaptive behaviors, including cyber aggression (Barlett et al., 2024; Goodboy & Martin, 2015; Moor & Anderson, 2019).

Previous studies have shown that moral disengagement and toxic online disinhibition mediate the relationship between Dark Triad traits and cyber aggression (Gholami et al., 2025; Kurek et al., 2019; Pabian & Vandebosch, 2023). However, longitudinal research examining the mechanisms through which these personality traits contribute to antisocial aggression remains limited (Sijtsema et al., 2019). Therefore, to address this gap, the present study employed a three-wave longitudinal design to investigate the mediating roles of moral disengagement and toxic online disinhibition in the association between Dark Triad traits and cyber aggression.

Dark Triad and Cyber Aggression

The Dark Triad is a personality framework encompassing three socially aversive traits: Machiavellianism, trait narcissism, and trait psychopathy (Paulhus & Williams, 2002). Specifically, Machiavellianism is characterized by manipulateness, a lack of empathy, and a strategic orientation to social interactions (Furnham et al., 2013). Narcissism is marked by grandiosity, entitlement, dominance, and an inflated sense of superiority (Maples et al., 2014; Paulhus & Williams, 2002). Psychopathy is associated with callousness, recklessness, and a tendency toward thrill-seeking behaviors (Jones & Paulhus, 2013).

The positive association between the Dark Triad and cyber aggression has been supported by both systematic reviews and meta-analyses (e.g., Barlett et al., 2024; Moor & Anderson, 2019). For instance, Barlett et al. (2024) conducted a meta-analysis of 211 studies and found a positive relationship between the Dark Triad traits and cyber aggression.

Compared with the majority of studies that employed cross-sectional designs, only a few have adopted longitudinal approaches to examine these associations over time (e.g., Fanti et al., 2012; Giumetti et al., 2022; C.-Y. Wang & Bi, 2025). For instance, C.-Y. Wang and Bi (2025) conducted a longitudinal study with Taiwanese social media users and found that narcissism, psychopathy, and Machiavellianism were positively associated with various forms of cyber aggression, including cyberstalking, exclusion, and outing. Nonetheless, findings across longitudinal studies have been mixed. For example, Fanti et al. (2012) examined a sample of 1,416 adolescents in Cyprus and reported that narcissism was not significantly related to cyber aggression (Fanti et al., 2012).

Moral Disengagement

Moral disengagement is a cognitive process that allows individuals to reinterpret unethical actions to appear acceptable, thereby reducing guilt, shame, and moral dissonance (Bandura, 2002b). Bandura (2002a) noted that moral disengagement is more likely to occur when individuals do not directly experience the consequences of their actions or when harm is psychologically or physically distant (Bandura, 2002a). This is especially relevant in online interactions, where the lack of social-emotional cues limits immediate feedback, making it easier to rationalize harmful behavior (Runions & Bak, 2015). For example, a perpetrator ridiculing a victim's profile picture on social media may not witness the victim's reaction, reducing their awareness of harm. This detachment further reinforces the perception that their actions are trivial or humorous.

Prior research suggests that individuals with Dark Triad traits are more likely to justify immoral behavior through mechanisms of moral disengagement (Sijtsema et al., 2019; Zhang et al., 2022). This tendency is particularly pronounced in individuals with high levels of psychopathy and Machiavellianism, as these traits emphasize self-interest and personal gain, often at the expense of others (Paulhus & Williams, 2002). Consistently, meta-analytic studies have found a strong positive correlation between moral disengagement and cyber aggression, further supporting the role of cognitive distortions in facilitating online misconduct (Barlett et al., 2024; L. Chen et al., 2017; Kowalski et al., 2014).

Toxic Online Disinhibition

In face-to-face interactions, individuals tend to regulate their behavior by suppressing impulsive reactions and avoiding offensive language, largely due to social norms and the fear of negative social evaluations or judgments (Joinson, 1998). However, in an online environment, these inhibitions may be weakened, leading some individuals to view the internet as relatively less monitored and regulated (Suler, 2004). According to online disinhibition theory, this perception increases the likelihood of engaging in behaviors they would typically avoid in real-world interactions (Cheung et al., 2020; Stuart & Scott, 2021; Suler, 2004). When these behaviors are positive (e.g., expressing oneself more openly or engaging in prosocial behaviors), they are referred to as benign online disinhibition. Conversely, when individuals feel emboldened to act aggressively, or believe that anonymity and reduced accountability shield them from consequences, this is referred to as toxic online disinhibition (Suler, 2004). For instance, in the case of toxic online disinhibition, an individual who avoids using derogatory language in face-to-face interactions to maintain their social image may feel empowered to do so online, where they perceive greater anonymity and reduced accountability for their actions.

Numerous studies highlight toxic online disinhibition as a significant risk factor for cyber aggression (Cheung et al., 2020; Huang et al., 2020; Kurek et al., 2019; Udris, 2014; Wu et al., 2023). Moreover, several studies have demonstrated that individuals with high levels of antisocial traits (e.g., Dark Triad traits) are more likely to hold toxic online disinhibition beliefs, further increasing the likelihood of engaging in aggressive online behaviors (Kurek et al., 2019; Wright et al., 2019; Wu et al., 2023).

Present Study

Although prior research has examined mediating mechanisms between the Dark Triad traits and cyber aggression, most studies have adopted cross-sectional designs (e.g., Gholami et al., 2025; Wu et al., 2023), which limit the capacity to capture dynamic processes over time. Conversely, although a few longitudinal studies have been conducted (e.g., Fanti et al., 2012; Giumetti et al., 2022; C.-Y. Wang & Bi, 2025), they often did not include mediators in their models, and their findings remain inconsistent—some have reported positive associations between the Dark Triad and cyber aggression (e.g., C.-Y. Wang & Bi, 2025), whereas others have found no significant relationships (e.g., Fanti et al., 2012). These discrepancies underscore the need for longitudinal research that incorporates potential mediating processes to better elucidate the temporal pathways linking the Dark Triad to cyber aggression.

To address this gap, the present study employs a longitudinal mediation model, collecting data at three time points to examine the mediating roles of moral disengagement (Time 2) and toxic online disinhibition (Time 2) in the relationship between the Dark Triad traits (Time 1) and cyber aggression (Time 3). This methodological approach provides stronger causal inference than previous cross-sectional studies. A key feature of this model is the inclusion of auto-regressive paths, which estimate the stability of variables across time points, and cross-lagged paths, which assess directional influences between variables (Little et al., 2015). By adopting a longitudinal design, this study aims to provide a more precise understanding of how Dark Triad traits contribute to cyber aggression and whether moral disengagement and toxic online disinhibition serve as critical mediators. The proposed structural model is presented in Figure 1.

Individuals with high levels of the Dark Triad traits—Machiavellianism, narcissism, and psychopathy—are more likely to justify immoral actions through moral disengagement (Sijtsema et al., 2019). Additionally, meta-analyses have consistently linked moral disengagement to cyber aggression (Barlett et al., 2024; L. Chen et al., 2017). Based on prior mediation studies showing that personality traits are related to cyber aggression through moral disengagement (Charalampous et al., 2020; Gholami et al., 2025; Pabian & Vandebosch, 2023; Wu et al., 2023; Zhang et al., 2022), we propose the following hypothesis:

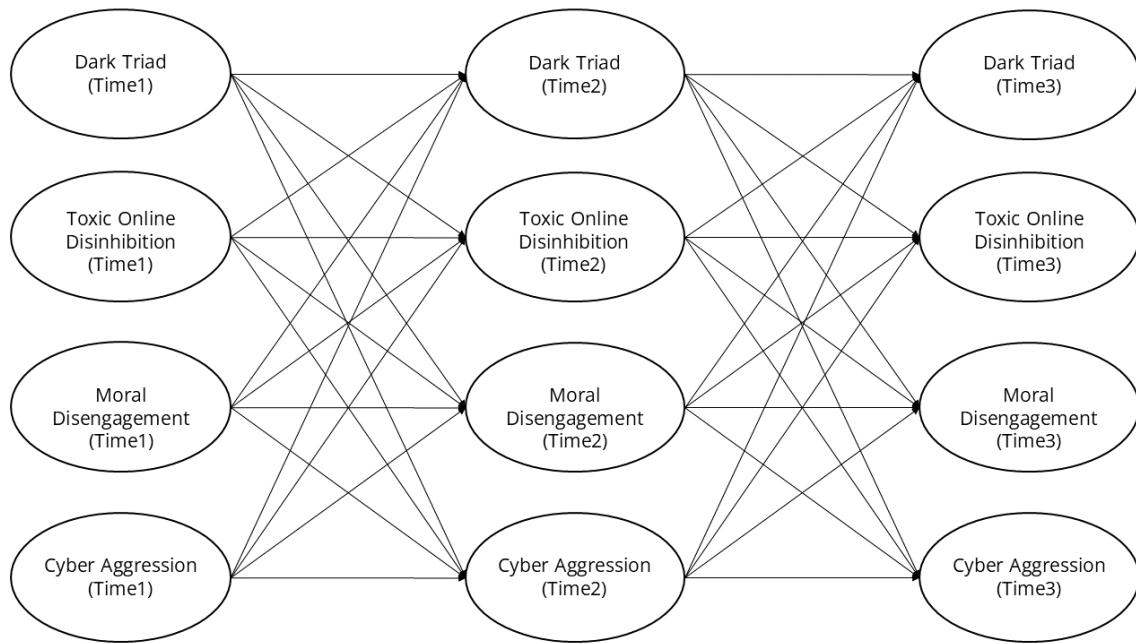
H1: Moral disengagement at Time 2 will mediate the relationship between the Dark Triad traits at Time 1 and cyber aggression at Time 3.

Similarly, toxic online disinhibition has been identified as a significant risk factor for cyber aggression (Cheung et al., 2020; Huang et al., 2020). Several studies indicate that individuals with high antisocial traits (e.g., Dark Triad traits) are more likely to endorse toxic online disinhibition beliefs (Wright et al., 2019; Wu et al., 2023), perceiving online environments as unregulated and anonymous (Kurek et al., 2019). This perception lowers social inhibitions

and increases the likelihood of engaging in cyber aggression. To further investigate this mediation effect, we propose the following hypothesis:

H2: Toxic online disinhibition at Time 2 will mediate the relationship between the Dark Triad traits at Time 1 and cyber aggression at Time 3.

Figure 1. *The Structural Diagram of the Multiple-Mediator Model.*



Note. Not all correlations and observation variables are displayed for ease of reading.

Methods

Participants and Procedure

This study received approval from the University's Research Ethics Committee for the Protection of Human Subjects. Taiwanese participants were recruited through social media platforms (e.g., Facebook, Gamer, and PTT) using a convenience sampling method. To encourage survey participation, a lottery system was implemented after each survey wave, where 25 randomly selected participants received a Starbucks gift voucher valued at 200 NTD. Data were collected at three time points, with a three-month interval between each wave, beginning in April 2022. Each survey remained open for one month, and reminder emails were sent 14 days later to encourage participation and completion.

The study was conducted via an online survey, which included a consent form and questionnaires. The consent form provided details about the purpose, participant rights, and the procedures. After providing informed consent, participants completed confidential questionnaires. Duplicate responses were identified and removed using email addresses to ensure data integrity. A total of 1,650 participants ($M_{age} = 27.06$, $SD_{age} = 7.14$, age range: 20 to 60; 941 females) completed the initial survey. Demographic characteristics of participants are presented in Table 1.

Participants received email reminders to complete the subsequent surveys. However, some participants were lost to follow-up. Specifically, 73 participants chose not to provide their email address after the first survey, and 700 participants did not respond to follow-up emails for the second survey. As a result, 877 participants completed the second survey ($M_{age} = 27.37$, $SD_{age} = 6.42$, age range: 20 to 58; 550 females).

For the third survey, all participants from the first wave were re-contacted, regardless of whether they had completed the second wave. A total of 817 participants completed the third survey ($M_{age} = 27.23$, $SD_{age} = 6.55$, age range: 20 to 60; 497 females). The majority had also participated in the second wave ($n = 625$, 76.4%), whereas a smaller subset had participated in the first wave but not the second ($n = 192$, 23.6%). In total, 625 participants completed all three waves of the survey ($M_{age} = 27.54$, $SD_{age} = 6.23$, age range: 20–58; 392 females).

Table 1. Demographic Characteristics of Participants.

Wave	Female	High school	Undergraduate	Graduate
1	941 (57.0%)	101 (6.1%)	1,092 (66.2%)	457 (27.7%)
2	550 (62.7%)	33 (3.8%)	577 (65.8%)	267 (30.4%)
3	497 (60.8%)	39 (4.8%)	536 (65.6%)	242 (29.6%)

To link responses across the three survey waves, participants voluntarily provided their email addresses, which were used exclusively for follow-up contact and matching responses between waves. The email information was stored separately from the survey data and deleted after data matching was completed. Participants were informed that their responses would remain confidential and untraceable to their identities, thereby reducing potential social desirability bias.

Little's MCAR test results indicated that missing data on the Dark Triad traits, moral disengagement, toxic online disinhibition, and cyber aggression were missing completely at random ($\chi^2_{(8)} = 14.179, p = .077$). Considering the substantial difference in participant numbers between the initial and second surveys, we conducted an analysis comparing those who dropped out with those who remained. The t-test results indicated that participants who completed the second survey did not significantly differ from those who dropped out in terms of Dark Triad traits, moral disengagement, toxic online disinhibition, and age at Time 1 ($p > .05$). However, their cyber aggression scores at Time 1 were significantly higher than those of participants who dropped out ($t = 4.062, p < .001, d = 0.197$). Nonetheless, given the trivial effect size, this significant difference is likely attributable to the large sample size rather than a meaningful practical distinction. Missing data were handled using the full information maximum likelihood (FIML) method, which maximizes the use of available data while accounting for missing values (Muthén & Muthén, 2017).

Measures

Cyber Aggression

This study assessed the frequency of cyber aggression by using the Traditional Chinese version of Cyber Aggression Scale (Chao, 2016). This scale comprises eight items based on various forms of cyberbullying as proposed by Willard (2007). An example item is *For someone I do not like, I will post or send messages that bother him or her*. Participants rated each item on a 5-point scale ranging from 1 (*never*) to 5 (*more than six times*). Higher scores indicate more frequent engagement in cyber aggression over the past six months. The scale exhibited high internal consistency, with Cronbach's α ranging from .942 to .948 across Time 1 to Time 3. Confirmatory factor analysis (CFA) conducted at Time 1 revealed that a one-factor model provided the best fit to the data, $\chi^2_{(20)} = 135.171, p < .001$, Comparative Fit Index (CFI) = .956, Root Mean Square Error of Approximation (RMSEA) = .059, Standardized Root Mean Square Residual (SRMR) = .025.

Dark Triad Traits

The Traditional Chinese version of Dark Triad scale, adapted from the Dirty Dozen (Jonason & Webster, 2010), measures narcissism with 3 items, Machiavellianism with 4 items, and psychopathy with 4 items. The Dark Triad comprises three distinct yet overlapping constructs (Paulhus & Williams, 2002). Considering these characteristics, many studies examine the reliability of both the overall Dirty Dozen scale and its individual subscales, including research on the development of the Dirty Dozen scale (e.g., Jonason & Webster, 2010; Sijtsema et al., 2019). A higher score indicates higher levels of the Dark Triad traits. Each item is rated on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is *I want everyone to admire me*.

Across three measurement time points, Cronbach's α values were as follows: overall Dark Triad (0.882–0.898), psychopathy (0.810–0.820), Machiavellianism (0.866–0.894), and narcissism (0.896–0.907). Following Jonason and Webster's (2010) finding that a three-factor model of the Dirty Dozen provided a better fit than a one-factor model, we adopted a three-factor model. At Time 1, the results of the CFA indicated that the three-factor model yielded acceptable fit indices ($\chi^2_{(41)} = 170.454, p < .001, CFI = .982, RMSEA = .044, SRMR = .030$).

Toxic Online Disinhibition

We assessed toxic online disinhibition using a 4-item subscale from the Online Disinhibition Scale developed by Udris (2014). In Udris's (2014) original study, the 12-item Online Disinhibition Scale was divided into two subscales: Benign Disinhibition and Toxic Disinhibition. Given this study's focus on maladaptive online behaviors—particularly those related to cyber aggression—we selected only the Toxic Disinhibition subscale (4 items), as it more directly captures disinhibited behaviors that are hostile, antisocial, or aggressive in nature. Participants rated their agreement with each item on a 5-point scale (1 = *strongly disagree* to 5 = *strongly agree*). A higher score reflects higher levels of toxic online disinhibition belief. An example item is *Writing insulting things online is not bullying*. The internal consistency of the scale, assessed using Cronbach's α , ranged from .877 to .909, reflecting good reliability. The CFA results at Time 1 showed a good model fit: $\chi^2_{(2)} = 12.089$, $p = .002$, CFI = .990, SRMR = .015, RMSEA = .055.

Moral Disengagement

The Traditional Chinese version of the Moral Disengagement Scale used in this study (Chao, 2016) was developed based on Bandura's (2002a) conceptual framework of moral disengagement. The scale consists of 17 items, with 3 items for moral justification and 2 items each for euphemistic labeling, advantageous comparison, displacement of responsibility, diffusion of responsibility, disregard or distortion of consequences, dehumanization, and attribution of blame. Participants rated their agreement with each item on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). A higher score indicates greater levels of moral disengagement. An example item is *Stealing to meet the needs of one's family is justifiable*. In this study, the internal consistency of the Moral Disengagement Scale demonstrated acceptable levels ($\alpha = .885 - .897$). The CFA results at Time 1 revealed that the 8-factor solution exhibited a good fit to the data ($\chi^2_{(91)} = 366.399$, $p < .001$, CFI = .964, SRMR = .028, RMSEA = .043).

Data Analysis

All analyses (e.g., longitudinal mediation model and measurement invariance) were conducted using Mplus version 7.4, with the exception of descriptive statistics (e.g., means and Pearson correlations), which were computed using SPSS version 29. Before conducting the longitudinal mediation model analyses, we followed the recommended procedure by Muthén and Muthén (2017) to test for measurement invariance. This process ensures that the measurement properties of the constructs remain consistent across the different periods, enabling a robust evaluation of the longitudinal relationships. Measurement invariance was examined across the three time points using a set of three nested models (i.e., configural, metric, and scalar), by progressively constraining the parameters of the measurement model. To compare the free estimation and constraint models, we used the difference in the comparative fit index (Δ CFI) and root mean square error of approximation (Δ RMSEA) rather than the chi-square difference test. The chi-square difference test is sensitive to even minor parameter changes in large samples, which makes Δ CFI and Δ RMSEA more appropriate metrics for our analysis (F. F. Chen, 2007; Luo et al., 2020). We considered the constrained models statistically equivalent than the free estimation models if the Δ CFI and Δ RMSEA were less than .01 (F. F. Chen, 2007). In this study, the measurement invariance testing confirmed that all scales demonstrated scalar invariance across the three time points.

Given the non-normality of the observed variables (see Table 2), a robust maximum likelihood estimator (MLR) was employed. Model fit was evaluated using the CFI ($\geq .90$), RMSEA ($\leq .07$), and SRMR ($< .08$; Hair et al., 2009). To capture the longitudinal nature of the data, correlations between error variables (i.e., unique factors) for each observation variable across time were included in the model (Liu et al., 2017).

Given the complexity of the longitudinal mediation model, we employed the isolated uniqueness parceling strategy. This strategy involves grouping similar items into distinct parcels, ensuring that each represents its unique characteristics (Lee & Whittaker, 2021). Specifically, we computed the mean score for each set of subscale items of the Dark Triad traits and formed three separate parcels, representing the unique characteristics of Machiavellianism, narcissism, and psychopathy. Similarly, we calculated the mean score for each set of subscale items of moral disengagement and grouped these mean scores into eight parcels. This parceling strategy aimed to simplify the model while capturing the essential elements of the Dark Triad and moral disengagement. Notably, this strategy has been employed in previous research using similar longitudinal designs (Sijtsema et al., 2019; Swart et al., 2011).

We conducted a power analysis using Monte Carlo simulations with the R package pwrSEM (Y. A. Wang & Rhemtulla, 2021), employing Spearman's correction (1904) to calculate structural parameters and setting factor loadings at 0.7. We focused on two indirect effects over time: (1) Dark Triad traits at Time 1 → Moral Disengagement at Time 2 → Cyber Aggression at Time 3, and (2) Dark Triad traits at Time 1 → Toxic Online Disinhibition at Time 2 → Cyber Aggression at Time 3. With a sample size of $N = 1650$ and 1000 simulation runs, the results indicated a power estimate of 1.0 for detecting these effects. Therefore, we conclude that the sample size of this study was adequate.

Results

Descriptive Statistics

Table 2 shows the means, standard deviations (*SD*), skewness, and Pearson correlations among all variables. Our primary focus was on the within-wave correlations to evaluate the stability of the variables over time. In this section, Dark Triad traits refer to the total score of the full Dark Triad scale, while Machiavellianism, narcissism, and psychopathy represent the total scores of the respective subscales. The overall Dark Triad score demonstrated strong test-retest reliability from Time 1 to Time 3 ($r = .711-.778, p < .001$). In terms of individual components, Machiavellianism showed acceptable reliability with correlations ranging from $r = .591$ to $.728$ ($p < .001$) between Time 1 and Time 3. Similarly, narcissism exhibited significant correlations across all three time points ($r = .601-.711, p < .001$). Furthermore, psychopathy demonstrated stable reliability over the three assessments, with correlations ranging from $r = .624$ to $.709$ ($p < .001$). These findings indicate that both the overall Dark Triad construct and its individual components are stable over time.

Similarly, the test-retest reliability of moral disengagement was also supported by significant correlations observed between Time 1, Time 2, and Time 3 ($r = .662-.744, p < .001$). Furthermore, acceptable test-retest reliability of toxic online disinhibition was observed between Time 1, Time 2, and Time 3 ($r = .691-.733, p < .001$). Lastly, the test-retest reliability of cyber aggression was acceptable, as indicated by the correlation between Time 1, Time 2, and Time 3 ($r = .734-.742, p < .01$). The results indicate that all variables are stable over time.

Table 2. Within- and Cross-Wave Pearson Correlations for Study Variables.

	DT ₁	M ₁	N ₁	P ₁	MD ₁	TD ₁	CA ₁	DT ₂	M ₂	N ₂	P ₂	MD ₂	TD ₂	CA ₂	DT ₃	M ₃	N ₃	P ₃	MD ₃	TD ₃	CA ₃	
DT ₁	—																					
M ₁	.886	—																				
N ₁	.765	.542	—																			
P ₁	.794	.589	.364	—																		
MD ₁	.578	.484	.333	.597	—																	
TD ₁	.597	.489	.320	.652	.606	—																
CA ₁	.561	.494	.313	.562	.610	.704	—															
DT ₂	.712	.662	.551	.568	.473	.462	.442	—														
M ₂	.559	.591	.394	.404	.356	.313	.322	.854	—													
N ₂	.477	.367	.601	.244	.252	.215	.213	.736	.492	—												
P ₂	.541	.462	.266	.624	.458	.490	.432	.807	.605	.342	—											
MD ₂	.451	.404	.249	.474	.662	.483	.478	.580	.470	.329	.594	—										
TD ₂	.558	.474	.317	.605	.505	.725	.636	.539	.370	.270	.600	.582	—									
CA ₂	.504	.448	.286	.526	.509	.617	.742	.518	.396	.284	.518	.548	.678	—								
DT ₃	.711	.617	.553	.574	.496	.459	.460	.778	.670	.579	.618	.495	.499	.480	—							
M ₃	.635	.640	.444	.460	.440	.364	.379	.714	.728	.444	.528	.440	.409	.420	.885	—						
N ₃	.550	.422	.646	.294	.284	.225	.269	.570	.402	.711	.272	.270	.246	.280	.768	.550	—					
P ₃	.542	.418	.269	.647	.478	.530	.473	.606	.473	.271	.709	.495	.569	.470	.777	.555	.354	—				
MD ₃	.434	.374	.219	.469	.672	.476	.434	.489	.393	.307	.487	.744	.471	.466	.574	.493	.334	.569	—			
TD ₃	.459	.371	.222	.535	.494	.691	.598	.479	.347	.223	.558	.529	.733	.632	.561	.433	.285	.654	.619	—		
CA ₃	.456	.391	.238	.488	.457	.599	.742	.452	.342	.237	.466	.431	.602	.734	.513	.407	.282	.564	.569	.744	—	
Mean	23.72	8.39	7.52	7.80	35.84	6.02	10.30	22.38	8.48	7.10	7.02	34.00	5.97	9.97	23.35	8.77	7.28	7.30	35.69	5.75	10.21	
SD	9.15	4.13	3.46	3.58	10.71	2.97	4.95	8.16	3.91	3.19	3.18	10.41	3.07	4.36	8.73	4.04	3.29	3.38	10.57	2.84	4.66	
Skewness	0.67	0.78	0.31	0.97	0.99	1.87	3.08	0.76	0.69	0.41	1.23	0.69	2.00	3.29	0.58	0.63	0.36	1.14	0.71	2.14	2.98	

Note. DT = Dark Triad; M = Machiavellianism; N = narcissism; P = psychopathy; MD = moral disengagement; TD = toxic online disinhibition; CA = cyber aggression. All correlations were significant at $p < .001$. The subscript number following the variable name indicates the time point of assessment (e.g., CA₁ is the total score of cyber aggression at Time1).

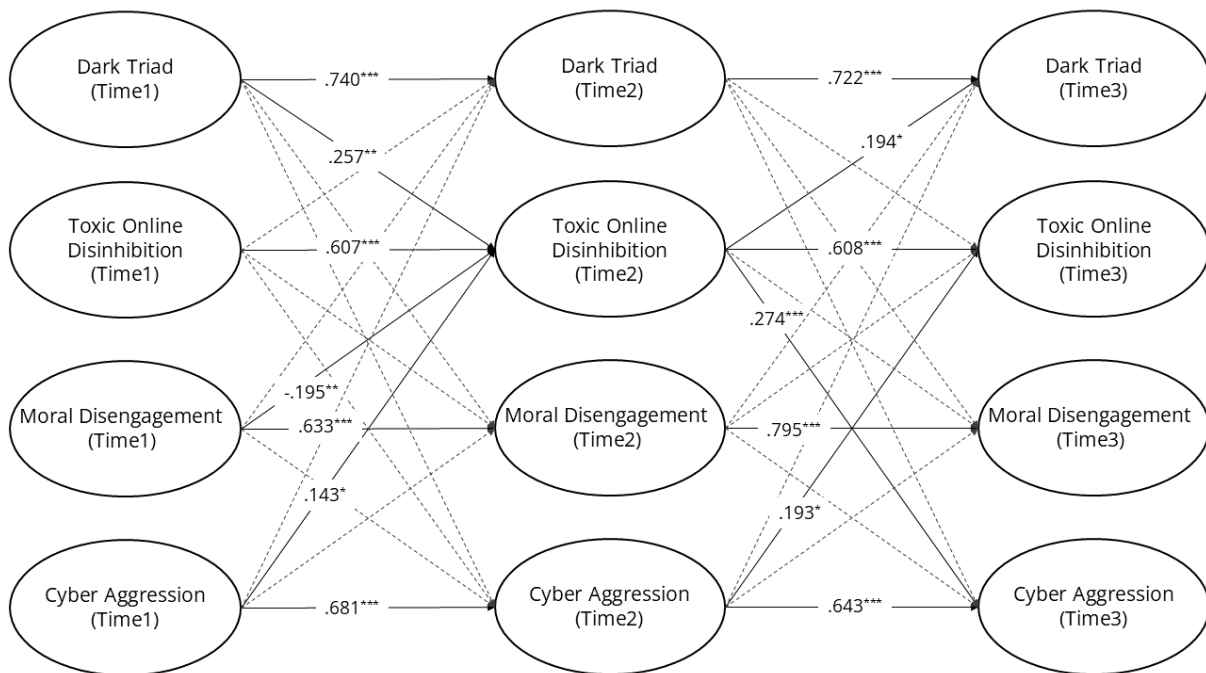
Longitudinal Mediation Model

We conducted a series of longitudinal multiple—mediator models to test the hypotheses. Analyses were first performed using the overall composite measure of the Dark Triad and then repeated at the subscale level (Machiavellianism, narcissism, and psychopathy). This approach allowed us to examine whether the mediating roles of toxic online disinhibition and moral disengagement differed across the three traits and to provide a more nuanced understanding of how each personality dimension contributes to cyber aggression.

Overall Dark Triad Model

The overall Dark Triad model exhibited acceptable model fit ($\chi^2_{(2294)} = 5407.270, p < .001, CFI = .924, RMSEA = .029, SRMR = .066$). After controlling for gender and age, toxic online disinhibition significantly mediated the relationship between the Dark Triad and cyber aggression (Indirect effect: 0.070, $p < .05$), whereas moral disengagement did not (Indirect effect = $-0.011, p = .278$). These findings support Hypothesis 2 but not Hypothesis 1. Figure 2 illustrates the longitudinal mediation model. We also examined reverse time—specific indirect paths (e.g., CA1 → TOD2 → DT3), which were found to be nonsignificant, suggesting that the relationship between cyber aggression and the Dark Triad is not mediated by toxic online disinhibition or moral disengagement in a reverse temporal sequence.

Figure 2. Illustration of Overall Dark Triad Model.

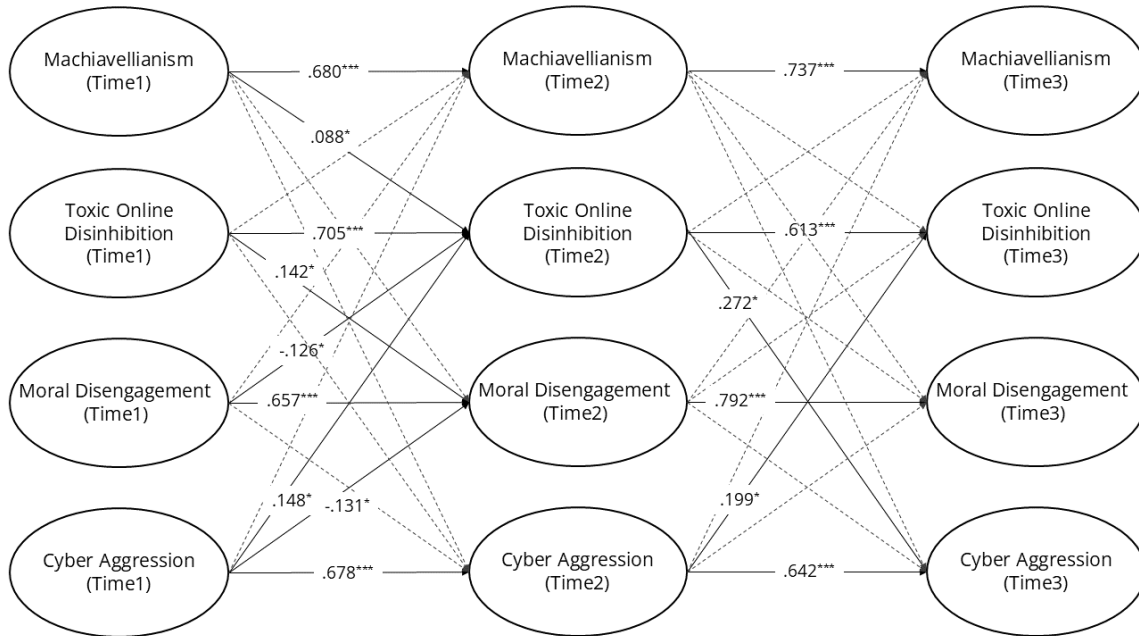


Note. Standardized path estimates are presented. Not all correlations and observation variables are displayed for ease of reading. Only significant paths are reported (* $p < .05$, ** $p < .01$, *** $p < .001$).

Machiavellianism Model

The Machiavellianism model yielded a good fit to the data, $\chi^2_{(2504)} = 5376.90, p < .001, CFI = .934, RMSEA = .026, SRMR = .064$. Within this model, toxic online disinhibition significantly mediated the relationship between Machiavellianism and cyber aggression (Indirect effect = 0.024, $p < .05$), whereas moral disengagement did not (Indirect effect = $-0.007, p = .200$). Figure 3 presents the structural model for the Machiavellianism pathway. The reverse time—specific indirect paths were not significant, indicating that toxic online disinhibition or moral disengagement measured at Time 2 did not explain the relationship between cyber aggression at Time1 and Machiavellianism at Time3.

Figure 3. Illustration of Machiavellianism Model.

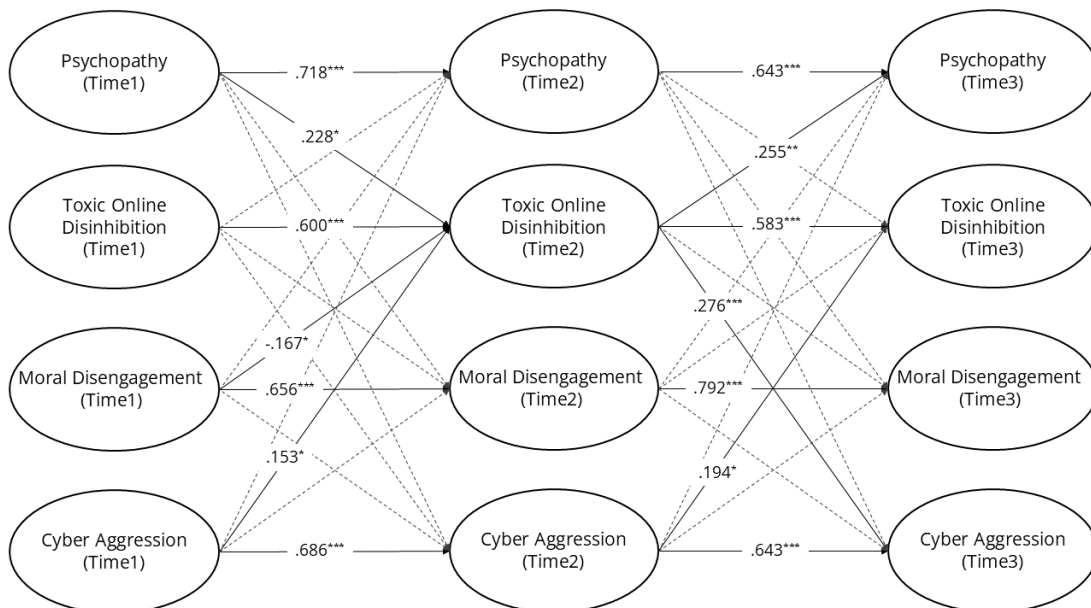


Note. Standardized path estimates are presented. Not all correlations and observation variables are displayed for ease of reading. Only significant paths are reported (* $p < .05$, ** $p < .01$, *** $p < .001$).

Psychopathy Model

The psychopathy model demonstrated a good fit to the data ($\chi^2_{(2504)} = 5572.821, p < .001, CFI = .927, RMSEA = .027, SRMR = .065$). In this model, toxic online disinhibition significantly mediated the relationship between psychopathy and cyber aggression (Indirect effect = 0.063, $p < .05$), whereas moral disengagement did not (Indirect effect = -0.004, $p = .922$). Figure 4 displays the structural model for the psychopathy pathway. The reverse time—specific indirect paths were also tested and found to be nonsignificant, indicating that neither toxic online disinhibition nor moral disengagement measured at Time 2 explained the relationship between earlier cyber aggression and later psychopathy.

Figure 4. Illustration of Psychopathy Model.

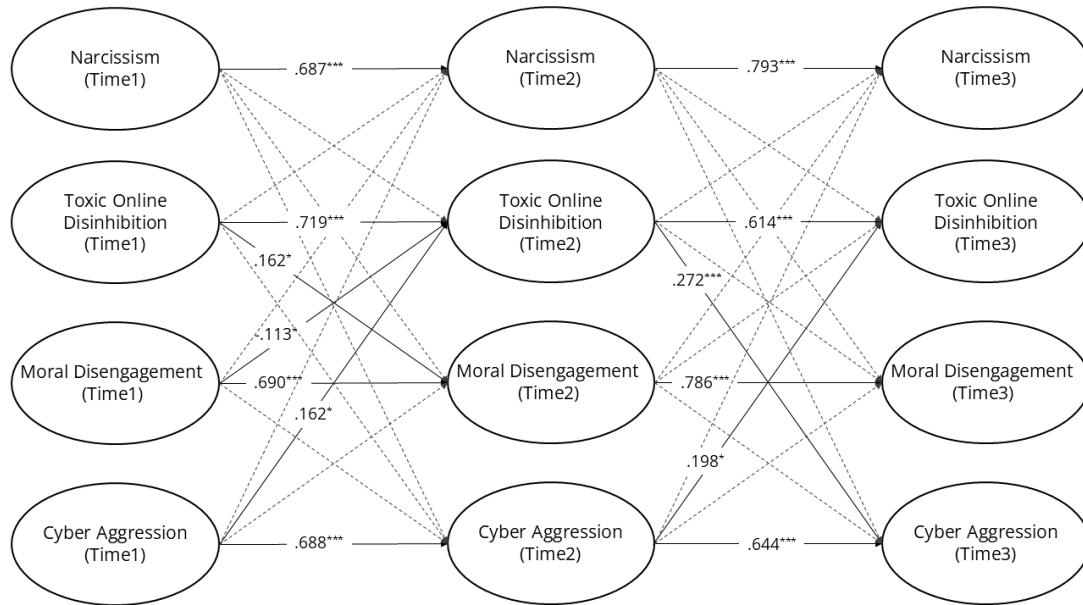


Note. Standardized path estimates are presented. Not all correlations and observation variables are displayed for ease of reading. Only significant paths are reported (* $p < .05$, ** $p < .01$, *** $p < .001$).

Narcissism Model

The narcissism model provided an acceptable model fit, $\chi^2_{(2294)} = 5120.433$, $p < .001$, CFI = .932, RMSEA = .027, SRMR = .064. In this model, neither toxic online disinhibition (Indirect effect = 0.011, $p = .207$) nor moral disengagement (Indirect effect = 0.003, $p = .273$) significantly mediated the relationship between narcissism and cyber aggression. Figure 5 illustrates the structural model for narcissism pathway. The reverse time—specific indirect paths were also tested and found to be nonsignificant, indicating that neither mediator explained the relationship from cyber aggression at Time 1 to narcissism at Time 3.

Figure 5. Illustration of Narcissism Model.



Note. Standardized path estimates are presented. Not all correlations and observation variables are displayed for ease of reading. Only significant paths are reported (* $p < .05$, ** $p < .01$, *** $p < .001$).

Discussion

This study investigated the mediating roles of moral disengagement and toxic online disinhibition in the relationship between the Dark Triad traits and cyber aggression using a longitudinal multiple—mediator model. Grounded in theoretical and empirical literature, we hypothesized that individuals with high levels of the Dark Triad traits would be more likely to exhibit moral disengagement and toxic online disinhibition. However, the results did not support moral disengagement as a significant mediator, regardless of whether the Dark Triad traits were treated as a composite construct or as individual components (i.e., psychopathy, narcissism, and Machiavellianism). In contrast, toxic online disinhibition emerged as a significant mediator linking the Dark Triad traits to cyber aggression. This mediating effect was evident both for the overall Dark Triad construct and for the specific traits of psychopathy and Machiavellianism. Narcissism, however, did not demonstrate a significant mediation pathway through toxic online disinhibition.

The Mediating Effect of Moral Disengagement

Previous cross—sectional studies have suggested that moral disengagement acts as a mediator between the Dark Triad traits and aggressive behaviors (Charalampous et al., 2020; Fang et al., 2020). However, our longitudinal mediation models—which provide a stronger test of causal processes over time (Swart et al., 2011)—did not support the mediating role of moral disengagement in linking Dark Triad traits to cyber aggression.

One possible explanation lies in the psychological profile of individuals with high levels of Dark Triad traits. As highlighted in prior research, such individuals often display emotional detachment, callousness, and a general disregard for moral norms (Paulhus & Williams, 2002). A recent survey further suggests that individuals with high levels of these traits may not experience cognitive dissonance when engaging in harmful behavior, and therefore may feel little need to engage in moral justification (Gajda et al., 2023). Their diminished concern for moral

standards or their reduced sensitivity to moral emotions may render moral disengagement irrelevant in the pathway to cyber aggression.

The Mediating Effect of Toxic Online Disinhibition

Our findings provide empirical support for the mediating role of toxic online disinhibition in linking Dark Triad traits to cyber aggression, consistent with prior research (Kurek et al., 2019), which indicates that individuals exhibiting high levels of Dark Triad traits are particularly inclined to endorse toxic disinhibition beliefs, viewing the internet as a space characterized by anonymity and lack of regulation. Such beliefs may lower social inhibitions and increase the likelihood of engaging in cyber aggression. To further clarify which traits drive this effect, we analyzed the three components of the Dark Triad traits—psychopathy, narcissism, and Machiavellianism—separately. Results showed that toxic online disinhibition significantly mediated the relationship between psychopathy and cyber aggression, as well as between Machiavellianism and cyber aggression, but not for narcissism.

In interpreting these results, Taiwan's cultural context may provide meaningful insight. Taiwanese society traditionally emphasizes social harmony, respect for norms, and emotional restraint (Kim & Markus, 1999). However, the anonymity and invisibility of online environments may weaken these cultural constraints, particularly for individuals already inclined to violate social expectations (C.-Y. Wang et al., 2025). For individuals with high levels of psychopathy or Machiavellianism—who often exhibit reduced empathy, emotional detachment, and weak internalized moral standards (Jones & Paulhus, 2013)—external social norms may typically function as behavioral regulators. In unregulated online spaces, these external cues are diminished, creating a setting in which toxic disinhibition beliefs may flourish. Consequently, individuals with high levels of psychopathy or Machiavellianism who strongly endorse toxic online disinhibition beliefs—perceiving cyberspace as lacking supervision and accountability—are more likely to act aggressively online. Although speculative, the cultural emphasis on social harmony in Taiwan may strengthen the mediating role of toxic online disinhibition.

In contrast, narcissism, often considered the “brighter” facet of the Dark Triad traits (Moor & Anderson, 2019; Rauthmann & Kolar, 2013), was not a significant mediator in this pathway. This may reflect the distinct motivational profile of individuals with high levels of narcissism, whose behaviors are primarily driven by self—enhancement and the pursuit of social admiration (Geng et al., 2015; Moor & Anderson, 2019; Rauthmann & Kolar, 2013). Individuals with high levels of narcissism may engage in cyber aggression as a means of defending or promoting their grandiose self—image, especially when their status or ego is threatened (Fanti & Henrich, 2015). Prior research also indicates that narcissism, especially when coupled with high self—esteem, is associated with relational aggression (Golmaryami & Barry, 2009). It is possible that self—esteem moderates the link between narcissism and cyber aggression. However, as this study did not assess self—esteem, we were unable to evaluate this potential interaction, which may explain the non—significant mediating effect.

Although previous research has highlighted toxic online disinhibition as a risk factor for cyber aggression among individuals with high levels of Dark Triad traits (Kurek et al., 2019), these studies have relied primarily on cross—sectional designs. To further investigate causal directionality, we explored an alternative mediation model, in which cyber aggression was hypothesized to influence Dark Triad traits through toxic online disinhibition. This exploratory model yielded no significant effects—whether tested for the overall Dark Triad or its individual traits—thereby reinforcing our primary finding that Machiavellianism and psychopathy serve as antecedents, toxic online disinhibition functions as a psychological pathway, and cyber aggression represents the behavioral outcome.

Limitations

The present study has some limitations that warrant consideration. First, our sample was limited to social media users in Taiwan, which may restrict the generalizability of our findings to other cultural contexts. In Taiwan, where social harmony and adherence to social norms are highly valued (Kim & Markus, 1999), individuals tend to be more behaviorally restrained in social interaction. This cultural context may accentuate the psychological mechanism of toxic online disinhibition. Specifically, when individuals in such contexts endorse this belief, they may perceive cyberspace as having less effective social supervision, thereby increasing their likelihood of engaging in cyber aggression. In this regard, the cultural context may have strengthened the mediating effect of toxic online disinhibition.

Second, the non—significant reverse time—specific indirect paths observed in this study—namely, the absence of evidence that cyber aggression predicts subsequent levels of Dark Triad traits via either toxic online disinhibition or moral disengagement—may be partly attributable to the relatively short assessment interval. Prior meta—analytic research indicates that personality traits typically require extended developmental periods to undergo meaningful change (Roberts et al., 2006). Accordingly, future research employing longer longitudinal timeframes may be better positioned to uncover the psychological mechanisms linking cyber aggression to the development of Dark Triad traits.

Third, all study measures relied on self—report questionnaires, which may be subject to biases such as social desirability (Van de Mortel, 2008). Although widely used in personality and social psychological research, self—report methods may not fully capture participants' actual behaviors. Future studies could incorporate multiple assessment methods (e.g., behavioral observations or digital trace data) to provide a more comprehensive understanding of the constructs examined.

Implications for Research and Practice

For researchers, addressing the limitations discussed previously may facilitate more in—depth investigations. Methodologically, extending the assessment intervals may help determine whether cyber aggression reflects enduring behavioral patterns rather than mere short—term fluctuations. Thematically, future research could examine whether cultural norms moderate the indirect relationship between the Dark Triad traits and cyber aggression. For instance, collectivist cultures (e.g., Taiwan), which emphasize social harmony and norm adherence, may yield different patterns than individualistic cultures (e.g., the United States), where autonomy and personal expression are more valued (Markus & Kitayama, 1991). Employing multi—group longitudinal multiple mediation models could facilitate direct cross—cultural comparisons of how moral disengagement and toxic online disinhibition function as mediators between Dark Triad traits and cyber aggression.

The absence of significant mediation effects for narcissism suggests that it may operate through alternative psychological pathways distinct from Machiavellianism and psychopathy. Prior evidence indicates that narcissism—particularly when coupled with high self—esteem—is linked to relational and indirect forms of aggression (Golmaryami & Barry, 2009), it is possible that cyber aggression among narcissistic individuals is more subtle and image—focused. Recent longitudinal findings have also pointed to potential associations between narcissism and indirect cyber aggression over time (C.-Y. Wang & Bi, 2025). To better understand these relationships, future research should examine self—esteem as a potential moderator and explore whether narcissism is more strongly associated with indirect rather than direct forms of cyber aggression.

For educators and mental health practitioners, the findings of this study offer valuable insights for preventing cyber aggression. Our research revealed that among individuals with high levels of the Dark Triad traits—particularly psychopathy and Machiavellianism, toxic online disinhibition is the primary psychological mechanism driving engagement in cyber aggression. This suggests that reducing toxic online disinhibition beliefs may be a promising strategy for preventing such behavior.

According to Udris (2014), toxic online disinhibition beliefs are often reflected in the following statements: “I don't mind writing insulting things about others online because it's anonymous.” “It is easy to write insulting things online because there are no repercussions.” “There are no rules online; therefore, you can do whatever you want.” “Writing insulting things online is not bullying.” To address these misconceptions, we recommend that educators and mental health professionals counter these beliefs through targeted instruction and group discussions. For example, the belief that online anonymity protects aggressors can be challenged by highlighting the traceability of digital footprints. Practitioners can explain that individuals who post insulting comments are not truly anonymous, as there are now various methods—such as IP address tracking—that can be used to identify them. This concept can be effectively illustrated by showing a news report demonstrating how law enforcement agencies use IP tracking to locate suspects.

Conclusion

Our study makes several significant contributions to the literature on cyber aggression. First, by employing a longitudinal mediation design, our research overcomes the limitations of previous cross—sectional studies and provides a more nuanced understanding of the time—sequenced relationships and dynamic interplay between

the Dark Triad traits and cyber aggression. Second, our findings reveal that toxic online disinhibition mediates the relationship between the Dark Triad traits—and specifically its dimensions of psychopathy and Machiavellianism—and cyber aggression, while moral disengagement does not play a significant mediating role. This insight advances our understanding of the psychological mechanisms underlying cyber aggression. Finally, our study offers potential implications for practice. For practitioners, the findings suggest that toxic online disinhibition may serve as a psychological mechanism underlying cyber aggression among individuals with elevated levels of Dark Triad traits, particularly Machiavellianism and psychopathy. Interventions that address toxic online disinhibition related beliefs—for example, through educational programs or group discussions aimed at critically examining online behavior norms—may hold promise in reducing the likelihood of cyber aggression.

Conflict of Interest

The authors have no conflicts of interest to declare.

Use of AI Services

ChatGPT was only used in the writing process to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Authors' Contribution

Cheng-Yen Wang: conceptualization, methodology, data curation, writing—original draft. **Yih-Lan Liu:** conceptualization, methodology, writing—review & editing. **Chia-Yun Chang:** writing—review & editing.

Acknowledgement

The authors are extremely grateful to all participants in our study. This work was supported by the National Science and Technology of the Republic of China under Contract No. MOST109—2511—H009—003—MY3.

Declaration of Publication Ethics

Informed consent was obtained from all individual participants included in the study. All procedures performed in studies involving human participants were in accordance with the ethical standards of Research Ethics Committee for Human Subject Protection, National Yang Ming Chiao Tung University (NYCU—REC—110—101).

References

- Bandura, A. (2002a). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, 31(2), 101–119. <https://doi.org/10.1080/0305724022014322>
- Bandura, A. (2002b). Social cognitive theory in cultural context. *Applied Psychology*, 51(2), 269–290. <https://doi.org/10.1111/1464-0597.00092>
- Barlett, C. P., Kowalski, R. M., & Wilson, A. M. (2024). Meta-analyses of the predictors and outcomes of cyberbullying perpetration and victimization while controlling for traditional bullying perpetration and victimization. *Aggression and Violent Behavior*, 74, Article 101886. <https://doi.org/10.1016/j.avb.2023.101886>
- Bauman, S., Cross, D., & Walker, J. (2013). *Principles of cyberbullying research: Definitions, measures, and methodology*. Routledge/Taylor & Francis Group.
- Chao, Y.-Y. (2016). *The relationships among narcissism, empathy, moral disengagement, and cyberbullying of college students* [Unpublished master dissertation]. National Yang Ming Chiao Tung University.

- Charalampous, K., Ioannou, M., Georgiou, S., & Stavrinides, P. (2020). Cyberbullying, psychopathic traits, moral disengagement, and school climate: The role of self-reported psychopathic levels and gender. *Educational Psychology, 41*(3), 282–301. <https://doi.org/10.1080/01443410.2020.1742874>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Chen, L., Ho, S. S., & Lwin, M. O. (2017). A meta-analysis of factors predicting cyberbullying perpetration and victimization: From the social cognitive and media effects approach. *New Media & Society, 19*(8), 1194–1213. <https://doi.org/10.1177/1461444816634037>
- Cheung, C. M. K., Wong, R. Y. M., & Chan, T. K. H. (2020). Online disinhibition: Conceptualization, measurement, and implications for online deviant behavior. *Industrial Management & Data Systems, 121*(1), 48–64. <https://doi.org/10.1108/IMDS-08-2020-0509>
- Eun Jahng, K. (2024). Factors influencing South Korean early adolescents' cyber aggression. *Children and Youth Services Review, 158*, Article 107483. <https://doi.org/10.1016/j.childyouth.2024.107483>
- Fang, J., Wang, X., Yuan, K.-H., Wen, Z., Yu, X., & Zhang, G. (2020). Callous-unemotional traits and cyberbullying perpetration: The mediating role of moral disengagement and the moderating role of empathy. *Personality and Individual Differences, 157*, Article 109829. <https://doi.org/10.1016/j.paid.2020.109829>
- Fanti, K. A., & Henrich, C. C. (2015). Effects of self-esteem and narcissism on bullying and victimization during early adolescence. *The Journal of Early Adolescence, 35*(1), 5–29. <https://doi.org/10.1177/0272431613519498>
- Fanti, K. A., Demetriou, A. G., & Hawa, V. V. (2012). A longitudinal study of cyberbullying: Examining risk and protective factors. *European Journal of Developmental Psychology, 9*(2), 168–181. <https://doi.org/10.1080/17405629.2011.643169>
- Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The dark triad of personality: A 10 year review. *Social and Personality Psychology Compass, 7*(3), 199–216. <https://doi.org/10.1111/spc3.12018>
- Gajda, A., Moroń, M., Królik, M., Małuch, M., & Mraczek, M. (2023). The dark tetrad, cybervictimization, and cyberbullying: The role of moral disengagement. *Current Psychology, 42*(27), 23413–23421. <https://doi.org/10.1007/s12144-022-03456-6>
- Geng, Y., Sun, Q., Huang, J., Zhu, Y., & Hand, X. (2015). Dirty dozen and short dark triad: A Chinese validation of two brief measures of the dark triad. *Chinese Journal of Clinical Psychology, 23*(2), 246–250.
- Gholami, M., Thornberg, R., Kabiri, S., & Yousefvand, S. (2025). From dark triad personality traits to digital harm: Mediating cyberbullying through online moral disengagement. *Deviant Behavior, 1*–19. <https://doi.org/10.1080/01639625.2025.2453445>
- Giumetti, G. W., Kowalski, R. M., & Feinn, R. S. (2022). Predictors and outcomes of cyberbullying among college students: A two wave study. *Aggressive Behavior, 48*(1), 40–54. <https://doi.org/10.1002/ab.21992>
- Golmaryami, F. N., & Barry, C. T. (2009). The associations of self-reported and peer-reported relational aggression with narcissism and self-esteem among adolescents in a residential setting. *Journal of Clinical Child & Adolescent Psychology, 39*(1), 128–133. <https://doi.org/10.1080/15374410903401203>
- Goodboy, A. K., & Martin, M. M. (2015). The personality profile of a cyberbully: Examining the Dark Triad. *Computers in Human Behavior, 49*, 1–4. <https://doi.org/10.1016/j.chb.2015.02.052>
- Huang, C. L., Zhang, S., & Yang, S. C. (2020). How students react to different cyberbullying events: Past experience, judgment, perceived seriousness, helping behavior and the effect of online disinhibition. *Computers in Human Behavior, 110*, Article 106338. <https://doi.org/10.1016/j.chb.2020.106338>
- Joinson, A. (1998). Causes and implications of disinhibited behavior on the Internet. In J. Gackenbach (Ed.), *Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications* (pp. 43–60). Academic Press.
- Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment, 22*(2), 420–432. <https://doi.org/10.1037/a0019265>
- 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. <https://doi.org/10.1177/1073191113514105>

- Kim, H., & Markus, H. R. (1999). Deviance or uniqueness, harmony or conformity? A cultural analysis. *Journal of Personality and Social Psychology*, 77(4), 785–800. <https://doi.org/10.1037/0022-3514.77.4.785>
- Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychol Bull*, 140(4), 1073–1137. <https://doi.org/10.1037/a0035618>
- Kurek, A., Jose, P. E., & Stuart, J. (2019). 'I did it for the LULZ': How the dark personality predicts online disinhibition and aggressive online behavior in adolescence. *Computers in Human Behavior*, 98, 31–40. <https://doi.org/10.1016/j.chb.2019.03.027>
- Lee, J., & Whittaker, T. A. (2021). The impact of Item parceling on structural parameter invariance in multi-group structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(5), 684–698. <https://doi.org/10.1080/10705511.2021.1890604>
- Little, T. D., Deboeck, P., & Wu, W. (2015). Longitudinal data analysis. In R. A. Scott, S. M. Kosslyn, & M. Buchmann (Eds.), *Emerging trends in the social and behavioral sciences* (pp. 1–17). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118900772.etrds0208>
- Liu, Y., Millsap, R. E., West, S. G., Tein, J., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486–506. <https://doi.org/10.1037/met0000075>
- Luo, J., Wang, M.-C., Ge, Y., Chen, W., & Xu, S. (2020). Longitudinal invariance analysis of the Short Grit Scale in Chinese young adults. *Frontiers in Psychology*, 11, Article 466. <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00466>
- Maples, J. L., Lamkin, J., & Miller, J. D. (2014). A test of two brief measures of the dark triad: The dirty dozen and short dark triad. *Psychological Assessment*, 26(1), 326–331. <https://doi.org/10.1037/a0035084>
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>
- Moor, L., & Anderson, J. R. (2019). A systematic literature review of the relationship between dark personality traits and antisocial online behaviours. *Personality and Individual Differences*, 144, 40–55. <https://doi.org/10.1016/j.paid.2019.02.027>
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8nd ed.). Muthén & Muthén.
- Pabian, S., & Vandebosch, H. (2023). The dark tetrad, online moral disengagement, and online aggression perpetration among adults. *Telematics and Informatics Reports*, 11, Article 100089. <https://doi.org/10.1016/j.teler.2023.100089>
- 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)
- Rauthmann, J., & Kolar, G. (2013). The perceived attractiveness and traits of the dark triad: Narcissists are perceived as hot, machiavellians and psychopaths not. *Personality and Individual Differences*, 54(5), 582–586. <https://doi.org/10.1016/j.PAID.2012.11.005>
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1–25. <https://doi.org/10.1037/0033-2909.132.1.1>
- Runions, K. C., & Bak, M. (2015). Online moral disengagement, cyberbullying, and cyber-aggression. *Cyberpsychology, Behavior, and Social Networking*, 18(7), 400–405. <https://doi.org/10.1089/cyber.2014.0670>
- Sijtsema, J. J., Garofalo, C., Jansen, K., & Klimstra, T. A. (2019). Disengaging from evil: Longitudinal associations between the dark triad, moral disengagement, and antisocial behavior in adolescence. *Journal of Abnormal Child Psychology*, 47(8), 1351–1365. <https://doi.org/10.1007/s10802-019-00519-4>
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101. <https://doi.org/10.2307/1412159>

- Stuart, J., & Scott, R. (2021). The Measure of Online Disinhibition (MOD): Assessing perceptions of reductions in restraint in the online environment. *Computers in Human Behavior*, 114, Article 106534. <https://doi.org/10.1016/j.chb.2020.106534>
- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology, Behavior, and Social Networking*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>
- Sun, L., Tian, X., & Zhu, W. (2024). The long-term effect of ostracism on cyber aggression: Mutually predictive mediators of hostile automatic thoughts and personal relative deprivation. *Current Psychology*, 43(30), 25038–25049. <https://doi.org/10.1007/s12144-024-06187-y>
- Swart, H., Hewstone, M., Christ, O., & Voci, A. (2011). Affective mediators of intergroup contact: A three-wave longitudinal study in South Africa. *Journal of Personality and Social Psychology*, 101(6), 1221–1238. <https://doi.org/10.1037/a0024450>
- Udris, R. (2014). Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale. *Computers in Human Behavior*, 41, 253–261. <https://doi.org/10.1016/j.chb.2014.09.036>
- Vale, A., Pereira, F., Gonçalves, M., & Matos, M. (2018). Cyber-aggression in adolescence and internet parenting styles: A study with victims, perpetrators and victim-perpetrators. *Children and Youth Services Review*, 93, 88–99. <https://doi.org/10.1016/j.chilyouth.2018.06.021>
- Van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *The Australian Journal of Advanced Nursing*, 25(4), 40–48. <https://ajan.com.au/index.php/AJAN/article/view/1817>
- Wang, C.-Y., & Bi, K. (2025). Exploring the influence of the dark triad on indirect cyber aggression: A longitudinal study of a taiwanese sample. *Cyberpsychology, Behavior, and Social Networking*, 28(2), 105–111. <https://doi.org/10.1089/cyber.2024.0303>
- Wang, C.-Y., Liu, Y.-L., & Chang, C.-Y. (2025). Investigating the effects of dark triad and anonymity on exclusionary cyber aggression: A social media experiment. *Cyberpsychology, Behavior, and Social Networking*, 28(8), 566–573. <https://doi.org/10.1089/cyber.2024.0577>
- Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science*, 4(1). <https://doi.org/10.1177/2515245920918253>
- Willard, N. E. (2007). *Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress*. Research Press.
- Wright, M. F., Harper, B. D., & Wachs, S. (2019). The associations between cyberbullying and callous-unemotional traits among adolescents: The moderating effect of online disinhibition. *Personality and Individual Differences*, 140, 41–45. <https://doi.org/10.1016/j.paid.2018.04.001>
- Wu, B., Xiao, Y., Zhou, L., Li, F., & Liu, M. (2023). Why individuals with psychopathy and moral disengagement are more likely to engage in online trolling? The online disinhibition effect. *Journal of Psychopathology and Behavioral Assessment*, 45(2), 322–332. <https://doi.org/10.1007/s10862-023-10028-w>
- Zhang, Z., Bian, S., Zhao, H., & Qi, C. (2022). Dark triad and cyber aggression among Chinese adolescents during COVID-19: A moderated mediation model. *Frontiers in Psychology*, 13, Article 1011123. <https://doi.org/10.3389/fpsyg.2022.1011123>

About Authors

Cheng-Yen Wang is an Assistant Professor at the Graduate Institute of Education, Tunghai University. He received his Ph.D. in Education from National Yang Ming Chiao Tung University. His research interests include cyberpsychology, cyberbullying, and internet addiction.

<https://orcid.org/0000-0003-0900-3905>

Yih-Lan Liu (PhD, University of Texas at Austin) is a professor of Educational Psychology in the Institute of Education at the National Yang Ming Chiao Tung University. Her current research focuses on bystander behavior in bullying, promotion of social emotional skills in anti—bullying, post—traumatic growth of victimization from bullying, and youth internet use.

<https://orcid.org/0000-0001-7721-3985>

Chia-Yun Chang is a PhD candidate in the Institute of Education at National Yang Ming Chiao Tung University, specializing in educational psychology. Her research interests include social media use and mental health, as well as the applications of digital technologies in psychological education.

✉ Correspondence to

Yih-Lan Liu, Institute of Education, National Yang Ming Chiao Tung University, Taiwan. No. 1001, Daxue Rd. East Dist., Hsinchu City 300093, Taiwan, elaineliu@nycu.edu.tw

© Author(s). The articles in Cyberpsychology: Journal of Psychosocial Research on Cyberspace are open access articles licensed under the terms of the [Creative Commons BY-SA 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/) which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited and that any derivatives are shared under the same license.