

From racial prejudice to cyber-racism perpetration: The effects of the dark tetrad, anonymity
perceptions, and the dimensions of online disinhibition

by

Jordan E. Scott

B.S., Iowa State University, 2021

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Psychological Sciences
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Major Professor
Dr. Christopher Barlett

Copyright

© Jordan E. Scott 2026.

Abstract

Cyber-racism is online communication that negatively targets another's race. This behavior has a negative impact on victim's mental health and there has been little research conducted on the perpetration of this act. The current research aimed to integrate prejudice theory and cyber-racism specific theory to gain a more encompassing understanding of cyber-racism perpetration. The two current studies further aimed to extend cyber-racism research by identifying personality variables that may explain individual difference in cyber-racism perpetration. Study 1 examined how the Dark Tetrad moderates the serial mediating influence of anonymity and online disinhibition in the relationship between racial prejudice and cyber-racism. Results suggest that anonymity and online disinhibition serially mediate the relationship between racial prejudice and cyber-racism perpetration, especially for those high in Dark Tetrad traits. Study 2 investigated whether specific aspects of online disinhibition impact the integrated cyber-racism model and how the Dark Tetrad moderate these relationships. Results suggest that online disinhibition may work more efficiently as a whole construct, as no subdimension of online disinhibition serially mediated the integrated cyber-racism model. This program of research provided important information on how individual differences and specific nuances of online disinhibition influence an integrated cyber-racism model as well as practical implications in terms of who is most likely to perpetrate cyber-racism and in what situations. By better understanding the situations and individual differences in which cyber-racism is most likely, we may be able to limit the negative impacts on victims and perpetrators.

Table of Contents

List of Figures	vi
List of Tables	vii
Chapter 1- Introduction.....	1
Prejudice Theory	2
Justification-Suppression Model.....	3
Cyber-Racism and Theory	5
Anonymity Perceptions.....	8
Online Disinhibition Effect.....	10
Individual Differences	12
The Dark Tetrad.....	12
<i>Narcissism</i>	12
<i>Machiavellianism</i>	13
<i>Psychopathy</i>	14
<i>Sadism</i>	15
<i>Dark Tetrad and Racial Prejudice</i>	16
Studies Overview	17
Chapter 2 - Study 1: Personality and Online Factors.....	18
Study 1 Hypotheses	19
Method	21
Participants and Procedure.....	21
Materials	22
Data Analysis Plan.....	24
Results.....	25
Correlations.....	25
Mediation	27
Moderated Mediation.....	28
Study 1 Summary.....	34
Chapter 3 - Study 2: Dimensions of Online Disinhibition.....	35
Study 2 Hypotheses	36

Method	39
Participants and Procedure.....	39
Materials	40
Data Analysis Plan.....	41
Results.....	41
Correlations.....	41
Mediation	44
Serial Mediation.....	45
Moderated Mediation.....	46
Chapter 4 - General Discussion	54
Integrated Cyber-Racism Theory.....	55
Individual Differences	56
Dimensions of Online Disinhibition.....	58
Limitations & Future Directions.....	59
Implications	61
Conclusion	62
References.....	63
Appendix A.....	78
Appendix B.....	79
Appendix C	80
Appendix D.....	81
Appendix E	82
Appendix F.....	83
Appendix G.....	84

List of Figures

Figure 1. Justification-Suppression Model	4
Figure 2. Cyber-Racism Theory and Support	8
Figure 3. Study 1 Conceptual Model	21
Figure 4. Low Dark Personality Path Model	29
Figure 5. High Dark Personality Path Model.....	29
Figure 6. Low Machiavellianism Path Model.....	30
Figure 7. High Machiavellianism Path Model.....	30
Figure 8. Low Narcissism Path Model	31
Figure 9. High Narcissism Path Model.....	31
Figure 10. Low Psychopathy Path Model.....	32
Figure 11. High Psychopathy Path Model	33
Figure 12. Low Sadism Path Model	34
Figure 13. High Sadism Path Model.....	34
Figure 14. Study 2 Conceptual Model	38
Figure 15. Low Dark Personality Path Model	47
Figure 16. High Dark Personality Path Model.....	48
Figure 17. Low Narcissism Path Model	49
Figure 18. High Narcissism Path Model.....	49
Figure 19. Low Machiavellianism Path Model.....	50
Figure 20. High Machiavellianism Path Model.....	51
Figure 21. Low Psychopathy Path Model.....	52
Figure 22. High Psychopathy Path Model	52
Figure 23. Low Sadism Path Model	53
Figure 24. High Sadism Path Model.....	54

List of Tables

Table 1. Study 1 Correlations	26
Table 2. Study 2 Correlations	43

Chapter 1- Introduction

Today, people spend an increased amount of time online. Findings from a recent study found individuals spend (i) over four hours a day on their phones (Kaylin, 2024), (ii) roughly 6 and a half hours on the Internet per day (Petrosyan, 2025), and (iii) about 2 hours a day on social media (Dixon, 2025). Through the increased accessibility and communication that the Internet offers, people have a vehicle to cause harm largely absent prior to the increased accessibility and affordability of the Internet. One such harm is cyber-racism, which is online communication that is harmful or degrading and directed at another person based on their race (Back, 2002). This widespread issue impacts a wide variety of individuals. Research has found that 25% of Black US adults (compared to 3% of White adults; Duggan, 2017) and 21% of Black US youth (aged 13-17) have been harassed online due to their race/ethnicity (compared to 4% of White youth; Vogels, 2022). Victims of cyber-racism face a number of deleterious mental health (e.g., depression, anxiety, and loneliness; Keum & Miller, 2017; Keum & Li, 2022) and physical health effects (e.g., increased alcohol consumption; Keum & Ahn, 2021).

Due to the negative outcomes afforded to victims of cyber-racism, it is important to explore the theoretical underpinnings that yield cyber-racism perpetration. Not only can such theoretical insights help guide intervention efforts, these gains can assist scholars in understanding why cyber-racism occurs and for whom cyber-racism is more likely. There is currently little research on the processes and variables that impact cyber-racism perpetration. The current work extended cyber-racism research by investigating the moderating role of the Dark Tetrad personality traits on an integrated theory used to predict cyber-racism. This programmatic research examined the relationships between several personality traits (e.g., narcissism, psychopathy, etc.), racial prejudice, anonymity, online disinhibition, and cyber-racism

perpetration. Drawing upon existing theory, I examined (i) the mediating role that anonymity and online disinhibition have on the relationships between racial prejudice and cyber-racism perpetration and whether the Dark Tetrad moderates that relationship (Studies 1 and 2), and (ii) what specific aspects of online disinhibition predict cyber-racism perpetration (Study 2). Identifying the processes and variables germane to cyber-racism perpetration is an important first step in enhancing our theoretical understanding of the processes and personality moderators germane to cyber-racism.

Prejudice Theory

The act of sending a racist online message or targeting someone online based on their race stems from base racial prejudices an individual holds. These genuine prejudices represent the underlying series of beliefs towards others that tend to be automatic via learned development over time (Devine, 1989). Many theories of prejudice posit that prejudice develops early in our lifespan and can be difficult to change (Nesdale, 2007). People learn early in their lives what is, and is not, socially and culturally acceptable to express regarding others. Indeed, prejudice beliefs are often hidden by individuals to align with one's social norms and, therefore, are often suppressed to an undetectable level unless situational factors allow for the expression of prejudice in a way that does not threaten their self-concept and social norm (Crandall & Eshleman, 2003). Established theory provides a broader explanation of how individuals mentally suppress and justify the expression of undesirable prejudice to satisfy societal expectations. The current research attempts to integrate several theories to better understand cyber-racism. Indeed, I believe that the best way to understand this "new" form of racism perpetration is to draw upon myriad theoretical frameworks that focus on processes related to (i) racial prejudice, (ii) cyber-racism, and (iii) online communication.

Justification-Suppression Model

To provide an encompassing approach of understanding the processes by which prejudice is expressed, Crandall and Eshleman (2003) developed the Justification-Suppression Model of Prejudice (JSM; Figure 1), which describes the processes and situations that influence whether prejudice is expressed or not. JSM posits that “genuine” negative prejudices, stemming from a variety of social, cultural, cognitive, and developmental factors, will lead to the expression of that prejudice in the absence of other situational or personality deterrents. However, if individuals are able to suppress their prejudices, then the likelihood of the expression of prejudice will likely be decreased. Indeed, research has shown several variables that can suppress one’s genuine prejudices, such as social norms, and personal standards, beliefs, and values (Crandall & Eshleman, 2005). Actively suppressing genuine prejudice requires sufficient cognitive effort to maintain (Crandall & Eshleman, 2003) and is, therefore, mentally taxing. As such, Crandall and Eshleman (2003) believed that one way to relieve the cognitive resources needed to suppress is to engage in justification, which creates an opportunity for the behavioral expression of prejudice to occur without feelings of guilt or shame because one’s genuine prejudice can be justified by the perpetrator. Beliefs, values, social situations, and cognition processes are all examples of possible justifications (Crandall & Eshleman, 2003). Therefore, JSM posits three instances in which prejudice is expressed: (i) when suppression is unengaged, (ii) when suppression is unsuccessful, or (iii) when suppression is engaged, but justification is enacted. Moreover, JSM posits that expressed prejudice will be unlikely when one’s genuine prejudices are suppressed without justification processes.

Figure 1

Justification-Suppression Model (Crandall & Eshleman, 2003)

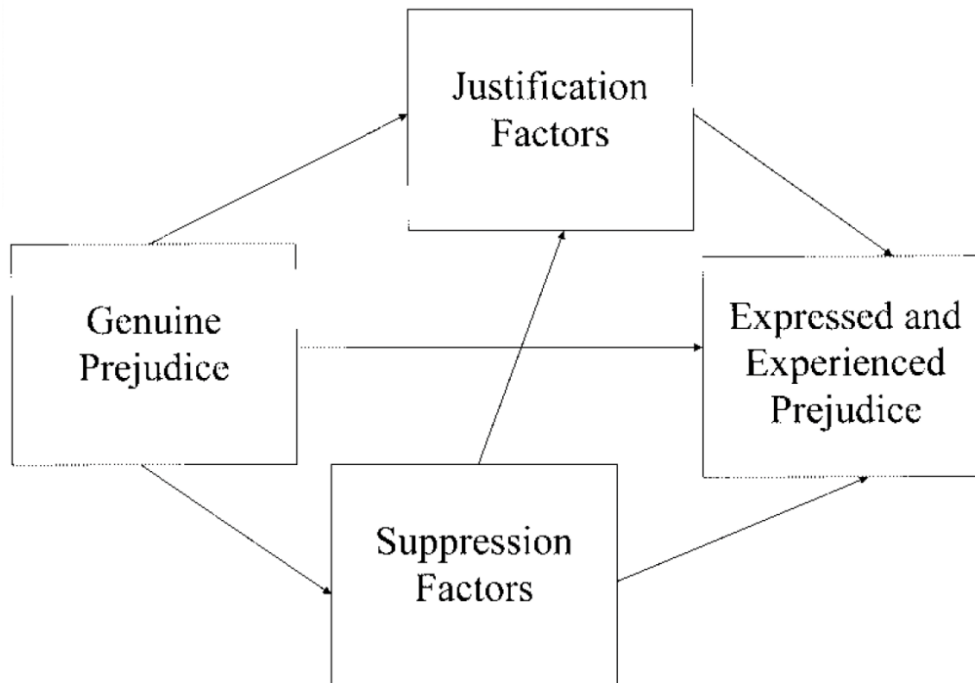


Figure 1. Justification-Suppression Model

The JSM was developed to be a model for general prejudice and is not specific to a type of prejudice. Research has supported the generalization of the JSM such that the model has been shown to explain expressions of prejudice towards a variety of groups. For example, research suggests that acceptance of homosexuality in tolerant locations may be due to religious justification (Hoffarth et al., 2018). Similar results have been found for other groups, such as Arab-Muslims (Arnoult et al., 2023), obese individuals (King et al., 2006), racial out-groups (White & Crandall, 2017), and the mentally ill (Seo et al., 2023). The current research focused on racial prejudice, specifically anti-black prejudice, and how the online world creates unique situational factors in which the expression of racial prejudice is more frequent.

We are unaware of any published study that applies JSM theorizing to cyber-racism; however, previous research suggests that JSM postulates should translate to the online space. For instance, Barlett and Scott (2023) showed that the correlation between cyber-racism and racial prejudice was significant ($r = .66$), suggesting that cyber-racism should also be predicted by

suppression and justification processes – a claim untested in the literature, but examined in the current research. However, the online environment differs in important ways that makes transitioning JSM tenets to predict online behavior ill-advised. Indeed, Lieberman and Schroeder (2020) suggested that the online sphere holds key differences in communication and behavior compared to offline settings, such as fewer nonverbal cues, greater anonymity, higher rates of connection and information dissemination. For the purposes of the current study, we will focus primarily on anonymity perceptions, defined as the perception one has that they are unidentifiable while online (Wright, 2013). Within the context of the JSM, anonymity may provide individuals with a way to release the suppression of their genuine prejudices, which would lead to higher rates of the expression of prejudice. In other words, if individuals perceive that their messages are not readily identifiable in an online space and, therefore, will not reap any negative consequences from others (e.g., family, friends, co-workers, law enforcement, etc.) because of said anonymity, then the motivation to suppress – and expound the cognitive energy to suppress – should be lower. According to JSM, this decrease (or release) of suppression ultimately increases the likelihood of expressions of prejudice. The current research aims to better understand cyber-racism perpetration through the JSM lens. This will be done by integrating racial prejudice into established cyber-racism theory which includes online specific perceptions such as anonymity and online disinhibition.

Cyber-Racism and Theory

While there is a paucity of research applying JSM postulates to predict cyber-racism, theoretical gains have been posited and supported to explain the processes of cyber-racism, albeit without influence of traditional racism theories. Keum and Miller (2018) developed, without empirical evidence, a theoretical model that identifies the psychological processes germane to

cyber-racism. This model identified three routes to cyber-racism, and each route starts with anonymity perceptions, predicting cyber-racism through myriad processes – each process is its own route. One of the three routes – the one that has received some empirical attention and used in the current research – focuses on how anonymity perceptions will indirectly predict cyber-racism through online disinhibition, defined as behaving differently, often antisocially, while online compared to real-world social interactions (Suler, 2004; 2005). Important to the Keum and Miller (2018) model is that anonymity perceptions and online disinhibition are two independent constructs. Past work (e.g., Scott et al., 2022; Wu et al., 2017) and theory (Suler, 2004) suggest that these two constructs should be positively correlated, and past research had supported this claim ($r = .30$; Wu et al., 2017, and $r = .54$; Scott et al., 2022). The distinction between anonymity and online disinhibition is directly theorized by Keum and Miller (2018) who posited that the theoretical temporal ordering of variables in their model is that online disinhibition is the mediator between anonymity (the predictor) and cyber-racism (the outcome). To ensure a clear distinction between the two constructs in the current research, I selected a measure of anonymity perceptions (Wright, 2013) whose items are specific to not being “caught” by authority figures for harmful online behaviors because one’s identity is often hidden online. Moreover, I used two measures of online disinhibition whose items reflected the belief that an individual acts differently online than in the real world – one measure assesses these beliefs holistically (Study 1) while the other assesses these beliefs with different factors representative of online disinhibition (Study 2). These scales were specifically chosen to avoid any conflation in our constructs, which is germane to our statistical assumptions. Overall, the Keum and Miller (2018) model posits that the perception of one’s anonymity leads to a disinhibited state while online, that leads to subsequent cyber-racism perpetration.

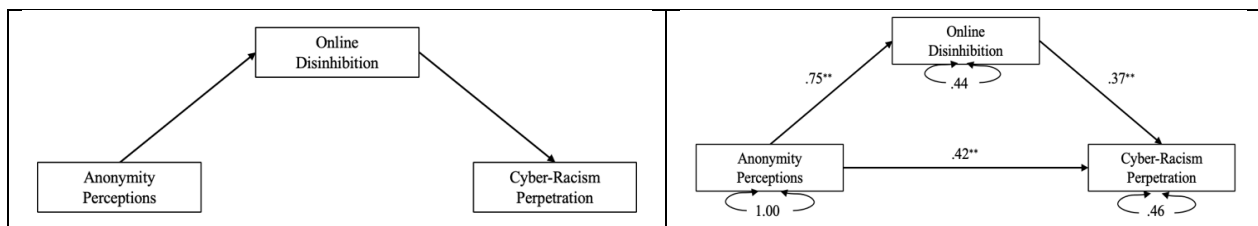
Evidence supporting the Keum and Miller (2018) model was found by Barlett and Scott (2023), who conducted six correlational studies. US adult participants completed measures of anonymity perceptions, online disinhibition, and cyber-racism, and results showed that online disinhibition mediated the relationship between anonymity and cyber-racism across all six studies. Moreover, a mini-meta-analysis of their six studies showed that online disinhibition partially mediated the relationship between anonymity and cyber-racism (Figure 2, right panel). In three of these studies, racial prejudice moderated the mediating effect of online disinhibition in the relationship between anonymity perceptions and cyber-racism, such that the relationship between online disinhibition and cyber-racism was stronger and significant for participants high (+1SD) , but not low (-1SD), on racial prejudice (Barlett & Scott, 2023) – suggesting that the Keum and Miller (2018) explicated mediated pathway is, in part, unique to the online space. Furthermore, Barlett and Scott (2023) also showed that online disinhibition mediated the relationship between anonymity perceptions and cyber-racism while controlling for cyberbullying perpetration – suggesting that the Keum and Miller (2018) posited mediated pathway is, in part, unique to the racism literature. Thus, cyber-racism combines aspects from prejudice and cyberbullying literatures to inform the current research.

Importantly, data from Barlett and Scott (2023) showed that the mediating effect of online disinhibition in the relationship between anonymity and cyber-racism, explicated by the Keum and Miller (2018) model, was, in part, unique to the online space by controlling for in-person aspects of the model tenets (e.g., in-person disinhibition), and the results showed that the modeled relationships remained significant. Overall, online racism is a behavior that draws on many literatures – racism, cyberbullying, online disinhibition – but represents a unique online behavior that juxtaposes these diverse topics. Therefore, the online space may provide an avenue

for individuals who hold genuine racial prejudices, but do not wish to be seen as undesirable in social setting, to expound such vitriol. Interestingly, such theorizing may even be seen in individuals who report low levels of racial prejudice, but have racial implicit biases or stereotypes, because there is now the opportunity for anonymous social interactions while online. Research has shown that individuals may even hold nonprejudiced self-concepts by believing others are more prejudice than themselves (Saucier, 2002). These ways of justifying one's prejudice (implicit or explicit) may disappear when the opportunity to express undesirable rhetoric is anonymous, thus releasing the suppression of said prejudice as current theory states.

Figure 2

Cyber-Racism Theory and Support (Left panel: Adapted Keum and Miller (2018) Model; Right panel: Barlett and Scott (2023) Meta-Analytic Results)



Anonymity Perceptions

Theoretically, it is my position that the integration of the JSM (Crandall & Eshleman, 2003) and Keum and Miller (2018) model is crucial for the understanding of cyber-racism perpetration. Anonymity perceptions, theoretically, is likely to be the crux of such integration, such that anonymity starts the Keum and Miller (2018) model (see Figure 2; left panel) and offers a release of suppression via JSM (Crandall & Eshleman, 2003). Moreover, Keum and Miller (2018) posit the relationship between anonymity perceptions (proxy for suppression in JSM for my research) and cyber-racism is mediated by online disinhibition. Therefore, it is

crucial to justify the inclusion of anonymity and online disinhibition theoretically to show further support of their importance.

One primary difference between in-person and online communication is the ability to hide or misrepresent oneself in online spaces. Indeed, Keum and Miller (2018) identified anonymity perceptions as the primary predictor to cyber-racism process. A multitude of research has found that anonymity perceptions positively predict various types of aggressive and racism, including cyber-aggression and cyber-racism. Theory posits that anonymity influences aggressive behavior in multiple ways. For example, the Social Identity Model of Deindividuation Effects (SIDE; Reicher et al., 1995) posits that anonymity perceptions lessen one's personal identity which in turn increases their reliance on group norms. These group norms vary across situations and may include a variety of antisocial norms which may increase the aggression one demonstrates online.

Generally, research has found that people in anonymous situations are more likely to act harmfully online versus those who are identifiable (Nitschinsk et al., 2022; Wright, 2014). People's anonymity perceptions may differ based on their experiences, the situations that they are in, and their motivations. Wright (2013) found that anonymity (avoiding detection from authority figure and avoiding retaliation from victim) were positively predictive of future cyber-aggression. This suggests that people's anonymity perceptions have similar impacts on negative online behavior even when they differ on motivations. These findings demonstrate the important role anonymity plays in the way people act online and emphasizes the theorized route that Keum and Miller (2018) proposed. Incorporating anonymity perceptions into both proposed studies is necessary to better understand the underlying processes of cyber-racism and how the Keum and Miller (2018) model fits within the JSM (Crandall & Eshleman, 2003).

Online Disinhibition Effect

The Keum and Miller (2018) model posits that online disinhibition mediates the relationship between anonymity perceptions and cyber-racism. As stated previously, behavior differs in the real world compared to the online world. Suler (2004; 2005) proposed the Online Disinhibition Effect, which suggests that the online world disinhibits an individual's behavior due to: (a) dissociative anonymity (ability to remove one's name/identity from their interactions), (b) invisibility (social interactions occurring when people cannot physically see one another), (c) asynchronicity (a temporal gap between when harm is perpetrated to when the victim receives the harm), (d) solipsistic introjection (the perceived merging of the real and online worlds), (e) dissociative imagination (the ability to psychologically escape or dissociate from online behaviors), and (f) minimization of status and authority (the absence or reduction of social cues that display status or authority). Not all six components are needed to behave disinhibited online, and differences in the importance of any of these six online disinhibition factors stems from one's personality, prior experience, and modalities of online communication (Suler, 2004). It has been suggested that different components play specific roles in how the online disinhibition is expressed, such that anonymity may be increasingly important for toxic online disinhibition compared to benign online disinhibition (Lapidot-Lefler & Barak, 2015).

As mentioned previously, Suler (2004) identified six components that could be the background reasons that online disinhibition takes place. A paucity of research has measured all six components of online disinhibition in a single study, which allows for statistical tests that examine what online disinhibition component is more, or less, important for the specific measured behavior. Unfortunately, the scant research in this domain is mixed. For instance, anonymity and invisibility (an online disinhibition component) have been found to predict online

threats – the number of threats and the likelihood of threatening another online decreased when participants were not anonymous and could be physically seen by the other person (Lapidot-Lefler & Barak, 2012). Similarly, anonymity and asynchrony (an online disinhibition component) were found to directly predict cyberbullying perpetration, but invisibility was not significantly related to cyberbullying (Wang & Ngai, 2020). Overall, anonymity and various online disinhibition components are important for nefarious online behavior, but the mixed findings call into question the veridical nature of the importance of online disinhibition. To address this issue, I will first examine the mediating role of online disinhibition in the relationship between anonymity perceptions and cyber-racism using a holistic measure of online disinhibition (Stuart & Scott, 2021). If significant, this would support the importance of online disinhibition in cyber-racism theorizing and replicate past research (Barlett & Scott, 2023). Then, in Study 2, I will use a measure of online disinhibition that segments online disinhibition into its six components (Cheung et al., 2020) to determine which, if any, components of online disinhibition are most important in cyber-racism. No published study that I am aware of has examined such relationships.

Individual Differences

While the primary objective of the current research was to integrate the JSM (Crandall & Eshleman, 2003), Online Disinhibition Effect (Suler, 2004; 2005), and the Keum and Miller (2018) model of cyber-racism, we are unaware of any research focusing on whether such theoretical tenets are ubiquitous or moderated. Therefore, in both studies I assessed the Dark Tetrad as potential moderators of the integrated model.

The Dark Tetrad

The Dark Tetrad is a set of personality traits (i.e., narcissism, Machiavellianism, psychopathy, and sadism) that are dispositioned toward amoral and antisocial behavior (Mededović & Petrovic, 2015). No published research has investigated how the Dark Tetrad (assessed as a whole [Dark Core] or as their individual constructs) moderates the theoretical underpinnings of cyber-racism. However, because cyber-racism is correlated with cyberbullying and traditional racism (Barlett & Scott, 2023), it is appropriate to draw upon that research to make informed hypotheses about the role of the Dark Tetrad on cyber-racism.

Narcissism

Narcissism can be categorized as feelings of entitlement, dominance, and superiority (Raskin & Hall, 1979). The tendency to exploit others for personal gains and difficulty learning from mistakes is another large part of narcissistic personality (Campbell et al., 2004, 2016). Narcissism has been found to be positively associated with a variety of aggressive online behaviors, such as cyberbullying (Alavi et al., 2002), cyber-trolling (Furian & March, 2023), and cyber intimate partner violence (Branson & March, 2021). Specifically, narcissism predicted cyberbullying perpetration in participants across different ethnicities, and especially for men (Brown et al., 2019). While there seems to be a direct link between narcissism and cyberbullying,

studies have additionally looked at possible mediators in this relationship. Fan et al. (2019) found that self-esteem mediated the relationship between narcissism and cyberbullying perpetration, such that those who scored high on narcissism had higher self-esteem which predicted lower levels of cyberbullying. This relationship can be seen in children when they engage in cyberbullying. A recent study found that narcissism indirectly predicted cyberbullying through attention-seeking when the children were classified as lonely (Wang et al., 2023). To receive the attention they desire, children may act out in negative ways, especially if they feel that they are lacking the deserved attention. It should be noted though that the relationship between narcissism and cyber-aggression is not always found. Goodboy and Martin (2015) did not find narcissism to be a unique predictor of cyberbullying. Though there appears to be conflicting evidence on how narcissism predicts cyberbullying, recent meta-analytic work found a significant positive relationship across roughly 45 studies with an effect size of .22 (Xu et al., 2024). It should be noted though that within these meta-analytic results, narcissism had the weakest relationship with cyberbullying compared to other members of the Dark Triad. As these relationships pertain to cyberbullying specifically, additional research is needed within the realm of cyber-racism to test whether similar relationships appear when race is the specific target.

Machiavellianism

The Machiavellianism trait can be described as a manipulative personality (Christie & Geis, 1970). There have been mixed results in determining Machiavellianism's role in cyber-aggression prediction. In a Malaysian sample, Machiavellianism was not found to be related to either cyberbullying or cyber-trolling (Alavi et al., 2022). Similarly, Gajda et al. (2022) found that even though Machiavellianism was significantly correlated with cyberbullying, it was not a significant predictor in the path model when controlling for the shared variance between the

Dark Tetrad traits. This lack of significant results extends to various other cyber-aggressive acts, such intimate partner cyberstalking. March et al. (2020) found that all three other Dark Tetrad constructs (narcissism, psychopathy, and sadism) were positive predictors in intimate partner stalking, but Machiavellianism was not.

Several studies though, have found that Machiavellianism is a predictor in cyber-aggressive behavior. In a longitudinal study, Machiavellianism positively predicted both cyberbullying perpetration and victimization (Giumetti et al., 2022). Brown et al. (2019) similarly found that all four Dark Tetrad traits were predictors of cyberbullying across multiple ethnicities. A recent study found that Machiavellianism was a stronger predictor of future cyberbullying perpetration for Chinese and Indian boys rather than girls (Wright et al., 2022). Though there appears to be a positive association/correlation between Machiavellianism and cyberbullying (Wright et al., 2022), limited studies find it to act as a predictor. It should be noted though that recent meta-analytic work found across 30 studies, Machiavellianism and cyberbullying had a significant positive relationship of .30 (Xu et al., 2024). These conflicting results suggest that Machiavellianism is not likely the strongest predictor of cyber-aggression but should still be examined to determine if additional context based on the type of aggression impacts the strength of Machiavellianism's predictability.

Psychopathy

The main tenets of psychopathy are high impulsivity and low empathy (Hare, 1985; Lilienfeld & Andrews, 1996). Compared to the previous facets of the Dark Tetrad, psychopathy has consistently been found to be a predictor of various forms of cyber-aggression. Pabian et al. (2015) found that psychopathy was the only predictor of cyber-aggression in adolescents, even while controlling for gender and age (note the date of publication was before sadism was

incorporated into the Dark Core). More specifically, psychopathy has been identified as a unique predictor of cyberbullying perpetration, such that higher psychopathy scores lead to more cyberbullying perpetration (Goodboy & Martin, 2015). Gibb and Devereux (2014) found similar results, such that men who scored higher on psychopathy engaged in more cyberbullying behavior. Indeed, meta-analytic work has found a significant positive relationship between psychopathy and cyberbullying across 30 studies with an effect size of .36 (Xu et al., 2024).

Research into psychopathy and cyber-aggression has additionally focused on possible mediators of this relationship. Sest and March (2017) found that lower empathy mediated the relationship between psychopathy and cyber-trolling. Thus, people with higher levels of psychopathy had lower levels of empathy which predicted higher amounts of cyber-trolling. Higher psychopathy scores have also been mediated by impulsivity in its relationship with cyberbullying perpetration (Park et al., 2024). Lower empathy and impulsivity acting as mediators is not surprising, as these are the key components of psychopathy in general (Hare, 1985; Lilienfeld & Andrews, 1996).

Sadism

Sadism is defined as enjoying physical or psychological harm of others (O'Meara et al., 2011). Often, provocation is needed to instigate aggression for those high in the other three Dark Tetrad traits. Provocation may not be necessary for those who score higher on sadism, as these individuals derive joy from causing other harm regardless of the current situation (Reidy et al., 2011). Research has supported this idea, as those higher in sadism were found to induce noise blasts on innocent individuals unprovoked (Buckels et al., 2013).

Sadism has been one of the most consistent predictors of different forms of cyber-aggression. Generally, sadism has been found to be directly predictive of problematic

smartphone use (Balta et al., 2019). Smartphones have opened the door for easier access to the internet and social media sites which could contribute to the cyber-aggression we see now. More specifically, March et al. (2020) found that sadism was a positive predictor of intimate partner cyber-stalking. Similarly, sadism has been found to be a predictor of anti-social sexual behavior on Tinder, such that those high in sadism were more likely to use Tinder to sexually coerce others and focus on their own sexual desires than their partner (Duncan & March, 2019).

Sadism as a predictor is additionally seen within the cyberbullying and cyber-trolling relationships. Van Geel et al. (2017) found that sadism was a predictor of both traditional bullying and cyberbullying, such that higher scores of sadism increased instances of in-person bullying and cyberbullying. Similar to results across the rest of the Dark Tetrad, sadism was found to predict cyberbullying perpetration across ethnicities, but it should be noted that scores on these constructs were higher for men (Brown et al., 2019). Results in cyber-trolling show similar relationships. March et al. (2024) found that self-reported cyber-trolls had higher levels of sadism compared to non-trolls. In another study, sadism was found to predict more cyber-trolling perpetration, especially in men (Furian & March, 2023). Additionally, March and Steele (2020) found that sadism predicted cyber-trolling, but only for those with high self-esteem. These results suggest that including sadism in the broader Dark Tetrad and research on cyber-aggression has provided important insights into what personality types are most likely to hurt others online.

Dark Tetrad and Racial Prejudice

In addition to correlating with cyber-aggression, individuals higher in the Dark Triad traits (i.e., narcissism, Machiavellianism, and psychopathy) hold higher prejudiced attitudes than participants low on the Dark Triad (Koehn et al., 2019). Similarly, Jonason, Underhill, and

Navarrate (2020) found that all Dark Triad traits were positively related to racial prejudice. Research has identified that the Dark Triad traits are not only related to prejudice generally but also, more specifically, with outgroup dehumanization (Capozza et al., 2019). However, the literature is largely mixed on which specific Dark Tetrad traits predict racial prejudice. Kocaturk and Bozdog (2020) found that only narcissism and psychopathy predicted xenophobia – characterized by hostility or fear towards foreigners – in college students. Other research showed that Machiavellianism and psychopathy predicted different manifestations of racism (i.e., classic vs. modern racism; Jones, 2013). More specific to the sadism trait, research suggests that people high in sadism are more likely to exhibit explicit prejudice (Greitemeyer, 2022). Acknowledging how the Dark Tetrad traits are related to racial prejudice is an important part of combining cyber-aggression literature and prejudice findings to investigate how the Dark Tetrad influence cyber-racism perpetration.

Studies Overview

To address the theoretical gaps in the literature, I conducted two cross-sectional studies that sampled US adults. Across both studies, the goal was to integrate the Justification Suppression Model (Crandall & Eshleman, 2003), Online Disinhibition (Suler, 2004), and cyber-racism theory (Keum & Miller, 2018) to test why and for whom cyber-racism is most likely. This program of research first tested the proposed integrated model in which anonymity and online disinhibition serially mediate the relationship between racial prejudice and cyber-racism in both studies. Second, the current studies examined whether the Dark Tetrad moderates the mediating influence of anonymity and online disinhibition in the relationship between racial prejudice and cyber-racism (Study 1) and whether specific aspects of online disinhibition impact these relationships (Study 2). This work further extended cyber-racism research to identify personality

variables that may explain individual difference in cyber-racism perpetration. The current work additionally integrated prejudice theory and cyber-racism theory to develop a clearer picture of the processes germane to perpetration. As the Keum and Miller (2018) model only looks at online disinhibition generally, the inclusion of the separate dimensions found within online disinhibition provided additional context into how personality and anonymity predict cyber-racism. No published research has focused on the moderating role of the Dark Tetrad on the integrated cyber-racism theory. Ultimately, the primary goal of this research was to examine how (a) the Dark Tetrad moderates the integrated prejudice and cyber-racism perpetration model and (b) the different dimensions of online disinhibition impact the Dark Tetrad's moderation.

Chapter 2 - Study 1: Personality and Online Factors

Cyber-racism negatively impacts the victims by increasing mental health problems (Keum & Miller, 2017). There has been little research examining the predictors and processes of cyber-racism perpetration and only one theory identifies the processes germane to cyber-racism specifically (Keum & Miller, 2018). However, because cyber-racism involves components of racism, we can rely on the underpinnings of the JSM (Crandall & Eshleman, 2003) to offer an alternative explanation of cyber-racism - genuine prejudice is often suppressed to adhere to social norms but can be expressed if proper justification is found. The unique online factors of anonymity and online disinhibition may act as a release of suppression, thus allowing for freer expression of prejudice. Study 1 tested this by testing the mediating influence of online disinhibition on the relationship between anonymity and cyber-racism, as suggested by Keum and Miller (2018). Study 1 furthered this understanding by testing the serial mediation of anonymity and online disinhibition in the relationship between racial prejudice and cyber-racism. Within this integrated model, anonymity and online disinhibition act as a release of the usual

suppression of prejudice which would then lead to an increased amount of expression (i.e., cyber-racism perpetration).

Moreover, in Study 1 I aimed to expand the current cyber-racism literature and theorizing by examining how individual differences in personality (i.e., the Dark Tetrad) moderate the mediating pathways in the integrated cyber-racism theory. Although the role of Dark Tetrad in cyber-racism prediction is too novel, we can rely on the more robust prejudice and cyberbullying research to make our predictions. Investigating how individual differences of personality influence the integrated theorized model furthered our understanding of who is most likely to be racist online.

Study 1 Hypotheses

The hypotheses for Study 1 pertain to the general relationships between the Dark Tetrad, racial prejudice, anonymity perceptions, online disinhibition, and cyber-racism perpetration. The first set of hypotheses will outline the anticipated correlations between the variables I measured.

H1: The Dark Tetrad traits will be positively correlated with racial prejudice, anonymity perceptions, online disinhibition, and cyber-racism.

H2: Racial prejudice will be positively correlated with anonymity, online disinhibition and cyber-racism.

H3: Anonymity will be positively correlated with online disinhibition and cyber-racism.

H4: Online disinhibition will be positively correlated with cyber-racism.

Next, the subsequent hypotheses¹ focus on our prediction regarding the mediating effects inherent in the juxtaposition of JSM and the Keum and Miller (2018) model that is depicted in Figure 2. Please note that Figure 3 is a conceptual model for ease of readability.

¹ For the following hypotheses, all predicted relationships will be positive and significant

H5: Online disinhibition will partially mediate the relationship between anonymity and cyber-racism, as theorized by Keum and Miller (2018).

H6: Anonymity and online disinhibition will partially serially mediate the relationship between racial prejudice and cyber-racism, as theorized within the integrated cyber-racism model in which anonymity and online disinhibition act as a release of suppression.

Finally, the next hypotheses² will predict the moderating role of the Dark Tetrad on the hypothesized serially mediated pathway from genuine prejudice to cyber-racism through anonymity perceptions and online disinhibition.

H7: The Dark Tetrad (as a composite of all four variables) will moderate the serial mediating pathway between racial prejudice and cyber-racism and the direct pathway of anonymity and cyber-racism as well as the direct pathway between racial prejudice and cyber-racism, such that these pathways will be significant for participants who are high on Dark Personality compared to participants who are low.

H8: Psychopathy will moderate the serial mediating pathway between racial prejudice and cyber-racism and the direct pathway of anonymity and cyber-racism, such that these pathways will be significant for participants high on psychopathy compared to participants who are low.

H9: Sadism will moderate the serial mediating pathway between racial prejudice to cyber-racism and the direct pathway of anonymity and cyber-racism, such that these pathways will be significant for participants high on sadism compared to participants who are low.

Figure 3
Study 1 Conceptual Model

² Due to the mixed results previously found regarding Machiavellianism and Narcissism's relationship with cyberbullying, no explicit hypotheses were drawn as there is not strong theoretical rationale. However, analyses were run including both these traits as moderators in an exploratory manner.

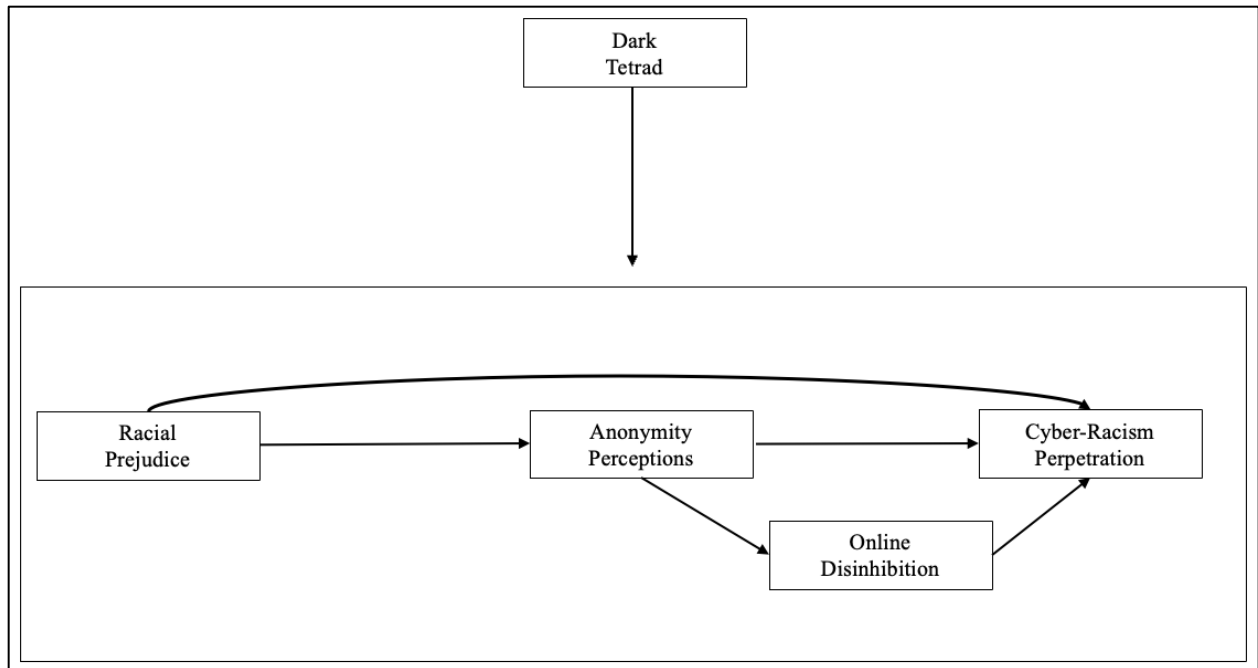


Figure 3. Study 1 Conceptual Model

Method

Participants and Procedure

According to an a priori power analysis (semPower; Moshagen & Bader, 2024), 469 participants were needed to achieve .80 power using the AGFI = .90 effect and when the model has five variables. To ensure sufficient data quality, attention checks were included in the survey. Failure of the attention checks resulted in the participant being dropped from analysis. To account for these potential drops, 619 participants were recruited from CloudResearch in exchange for \$2.00 (US) each. The inclusionary criteria required participants to be 18 years or older and be in the US.

Of the 619 participants who completed the survey, several were omitted from analysis due to failure of the reCAPTCHA element ($n = 42$), failure of completing any survey items ($n = 10$), and failure of one or more attention checks ($n = 22$). These data cleaning procedures resulted in 545 participants for data analysis. On average, participants were 39.31 years old ($SD = 12.05$)

and majority White (69%). Regarding sex, 51.7% identified as men and 47% as women. Overall, we remained substantially powered.

On Qualtrics, participants completed an informed consent and a series of surveys assessing several constructs, including demographic information, the Dark Tetrad, anonymity perception, online disinhibition, social desirability, racial prejudice, and cyber-racism perpetration. Three attention checks were included to ensure quality data collection. Lastly, participants were debriefed, thanked, and compensated.

Materials

Short Dark Tetrad (SD4). Participants completed the 28-item Short Dark Tetrad scale (SD4; Paulhus et al., 2021) to measure their levels of different dark personality components; Machiavellianism, narcissism, psychopathy, and sadism (See Appendix A). The SD4 uses a 1 (*strongly disagree*) to 5 (*strongly agree*) Likert-type scale in which participants rate their agreement with the items. During scale development (Paulhus et al., 2021), the SD4 demonstrated psychometric properties (e.g., construct validity, reliability, internal consistency, and clear distinction between the personality subscales). The SD4 is my preferred measure due to its inclusion of sadism as a dark personality trait, as this is a newer idea in personality research.

The SD4 included four subscales with seven items each. Each subscale assesses a separate personality trait, including Machiavellianism, $\alpha = .99$ (e.g., *Avoid direct conflict with other because they may be useful in the future*), narcissism, $\alpha = .98$ (e.g., *Group activities tend to be dull without me*), psychopathy, $\alpha = .98$ (e.g., *I tend to fight against authorities and their rules*), and sadism, $\alpha = .98$ (e.g., *Some people deserve to suffer*). For scoring participants' responses, I created composites for each of the subscales and a summed total score, given that

these were designed to identify specific personality traits and a combined dark personality score (Paulhus et al., 2021), and the reliability was acceptable for the composite ($\alpha = .94$).

Anonymity Perceptions. To measure participant's anonymity perceptions online, the 3-item Anonymity Perceptions Scale (Wright, 2013; See Appendix B) was used. This scale asks participants to rate their level of agreement with the items on a 1 (*strongly disagree*) to 5 (*strongly agree*) rating scale. This measure has correlated with online disinhibition, cyber-racism, cyberbullying, and traditional racism, which suggests this is a valid measure of anonymity perceptions (Barlett & Scott, 2023). A sample item includes, "I am confident that I would not be caught if I engaged in mean online behaviors.", and the reliability was acceptable ($\alpha = .91$).

Online Disinhibition. To measure the degree to which participants feel or behave differently in online interactions, the 12-item Measure of Online Disinhibition (MOD; Stuart & Scott, 2021; See Appendix C) was used. During scale development (Stuart & Scott, 2021), the MOD demonstrated psychometric properties (e.g., construct validity, reliability, and internal consistency). This measure asks participants rate their agreement on a 1 (*not at all like me*) to 5 (*very like me*) Likert-type scale. A sample item includes, "I act tougher on the internet than I do face-to-face.", and the reliability was acceptable ($\alpha = .94$).

Social Desirability.³ The Social Desirability Scale – 17 (SDS-17; Stober, 2001; See Appendix D) was used to measure social desirability. The original scale consisted of 17 items and participants are asked to respond true (coded 1) or false (coded 0) to each item. Stober (2001) explicated the removal of item 4 ("I have tried illegal drugs (for example, marijuana, cocaine, etc.)) from the total score. Across four studies, the SDS-17 (Stober, 2001) demonstrated

³ Social desirability was a covariate in the analyses due to the sensitivity of cyber-racism variable and items, which may cause participants to respond in a socially desirable manner (Grimm, 2010).

psychometric properties (e.g., construct validity, reliability, and internal consistency). A sample item includes, “In traffic I am always polite and considerate of others.”, and the reliability was acceptable ($\alpha = .81$).

Racial Prejudice. The Attitudes Towards Black People Scale (ATB; Brigham, 1993; See Appendix E) was used to assess levels of racial prejudice. This is a 20-item questionnaire that has participants rate their level of agreement with the items on a 1 (*strongly disagree*) to 9 (*strongly agree*) rating scale. Brigham (1993) found that the ATB demonstrated sound psychometric properties (e.g., construct validity, reliability, and internal consistency). A sample item includes, “I think that Black people look more similar to each other than White people do.”, and the reliability was acceptable ($\alpha = .94$).

Cyber-Racism Perpetration. Participants completed the Online Racism Perpetration Scale (Barlett & Scott, 2023; See Appendix F) to measure how often they engage in certain online behaviors in the past six months. This five-item questionnaire asks participants to indicate how often they engaged in various behaviors in the past six months using a 1 (*never*) to 5 (*all the time*) rating scale. Reliability and validity were demonstrated across seven studies for the Online Racism Perpetration scale (Barlett & Scott, 2023). A sample item includes, “Sent a racist meme (e.g., racist catchphrases, captioned photos, #hashtags etc.)”, and the reliability was acceptable ($\alpha = .95$).

Data Analysis Plan

All variables were centered before inclusion in the model, and one covariate was included: social desirability. For Study 1, correlations were calculated using a Spearman correlation, as the data was non-normal. The serial mediation was conducted to assess the base proposed integrated model of cyber-racism using PROCESS (Hayes, 2022). Finally, a moderated

mediation analysis was conducted using the lavaan package in R (Rosseel, 2012) to assess for changes in the mediation relationships based on the Dark Tetrad traits. For the moderated mediation analyses, a multi-group path analysis was conducted using a median split⁴ (i.e., High vs. Low) of the Dark Tetrad moderators. A MLR/Robust correction (Rosseel, 2012) was used to estimate model fit as the data was non-normal and bootstrapping was used to normalize the data.

Results

Correlations

Table 1 displays the zero-order correlations and descriptive statistics between the continuous variables. In line with H1, all Dark Tetrad variables were significantly positively correlated with racial prejudice, anonymity, online disinhibition, and cyber-racism. H2 was supported as racial prejudice was positively correlated with anonymity, online disinhibition, and cyber-racism. Similarly, H3 and H4 were supported such that, both anonymity and online disinhibition were positively correlated with each other and with cyber-racism.

⁴ I acknowledge that median splits hold several large limitations, but a multi-group path analysis approach was used for parsimony and to extract meaningful results from complex models. Prior to a median split approach, attempts were made to split the data by +/- 1SD but due to a floor effect and no variance in cyber-racism perpetration, the model would not converge.

Table 1*Study 1 Correlations and Descriptive Statistics*

	1	2	3	4	5	6	7	8	9
1: Anonymity Perceptions	----								
2: Online Disinhibition	.38**	----							
3: Cyber-Racism Perpetration	.31**	.29**	----						
4: Racial Prejudice	.32**	.29**	.33**	----					
5: Machiavellianism	.31**	.32**	.20**	.29**	----				
6: Narcissism	.31**	.12**	.22**	.20**	.38**	----			
7: Psychopathy	.46**	.28**	.29**	.31**	.29**	.46**	----		
8: Sadism	.48**	.39**	.30**	.37**	.45**	.33**	.58**	----	
9: Social Desirability	-.08	-.18**	-.05	-.10*	-.07	.17**	-.12**	-.21**	----
<i>M</i>	6.86	29.74	6.19	55.87	23.81	19.55	13.86	16.58	8.74
<i>SD</i>	3.29	12.68	3.28	29.08	5.06	6.10	5.86	6.48	4.09

Note: ** $p < .01$, * $p < .05$

Table 1. Study 1 Correlations

Mediation

For the purposes of replication, the Keum and Miller (2018) model was tested to ensure that the primary processes germane to our integrated model have support. In this analysis, anonymity was the independent variable, online disinhibition was the mediator, and cyber-racism was the dependent variable as stated in H5. It should be noted that all paths were estimated, which resulted in no degrees of freedom, and, therefore, PROCESS (Hayes, 2022) was used to estimate the model instead of the *laavan* package in R. Overall, results supported the Keum and Miller (2018) model and supported H5. In terms of the individual paths, anonymity perceptions were significantly associated with online disinhibition ($B = 1.42$, $SE = 0.16$, $p < .001$). Additionally, both anonymity perceptions ($B = 0.16$, $SE = 0.03$, $p < .001$) and online disinhibition ($B = 0.04$, $SE = 0.01$, $p < .001$) were significantly associated with cyber-racism. In support of H5, the indirect effect of online disinhibition was significant (*Indirect* $B = 0.06$ [.03,.08]). In short, anonymity perceptions significantly predicted online disinhibition to predict subsequent cyber-racism perpetration.

To test the integrated cyber-racism model (specified in H6), a serial mediation path analysis was run in PROCESS (Hayes, 2022) in which racial prejudice was the independent variable, anonymity and online disinhibition were the two serial mediators, and cyber-racism was the dependent variable. Results showed that racial prejudice was significantly associated with anonymity perceptions ($B = 0.04$, $SE = 0.004$, $p < .001$). Anonymity perceptions were significantly associated with online disinhibition ($B = 1.15$, $SE = 0.17$, $p < .001$). Additionally, racial prejudice ($B = 0.02$, $SE = 0.003$, $p < .001$), anonymity perceptions ($B = 0.11$, $SE = 0.03$, $p < .001$) and online disinhibition ($B = 0.03$, $SE = 0.01$, $p < .001$) were significantly associated with cyber-racism. In support of H6, the serial indirect effect of anonymity perceptions and

online disinhibition was significant (*Indirect B* = 0.0012 [.0006, .0022]). In other words, racial prejudice significantly predicted anonymity perceptions which in turn predicted online disinhibition, which predicted subsequent cyber-racism perpetration.

Moderated Mediation

To test H7, the moderated serial mediation path analysis in which the total dark personality⁵ score was the moderator on the previously described integrated cyber-racism model, the *laavan* package in R was used. Results should be interpreted with caution due to some components of model fit not meeting thresholds (Hu & Bentler, 1999), $\chi^2 = 9.51$ (2), $p = .009$, CFI = .96, TLI = .62, RMSEA = .12 [.05, .21], SRMR = .03. The overall dark personality construct moderated several direct paths but did not moderate the serial indirect path as hypothesized in H7. Anonymity perceptions were significantly associated with cyber-racism for participants high, $B = 0.15$ [0.06, 0.24] but not low, $B = 0.03$ [-0.00, 0.08] in dark personalities. Additionally, online disinhibition was significantly associated with cyber-racism for those high, $B = 0.05$ [0.03, 0.07] but not low, $B = 0.001$ [-0.00, 0.01] in dark personalities. These results partially support H7 as the overall dark personality construct moderated the direct path between anonymity perceptions and cyber-racism but did not moderate the direct path between racial prejudice and cyber-racism (see Figure 4/5).

Figure 4

Low Dark Personality Path Model

⁵ The total dark personality and subsequent subscales analyses were run separately due to too low power.

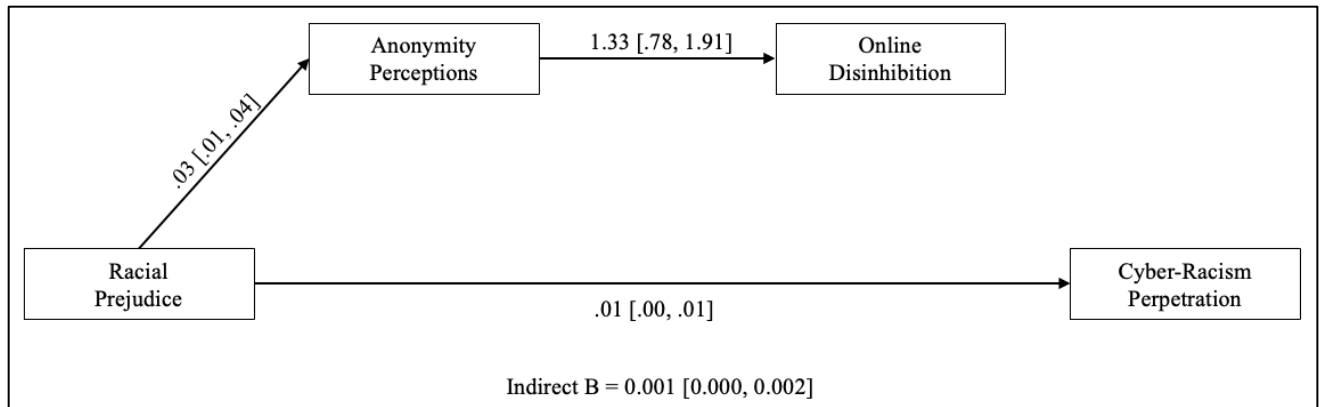


Figure 4. Low Dark Personality Path Model

Figure 5

High Dark Personality Path Model

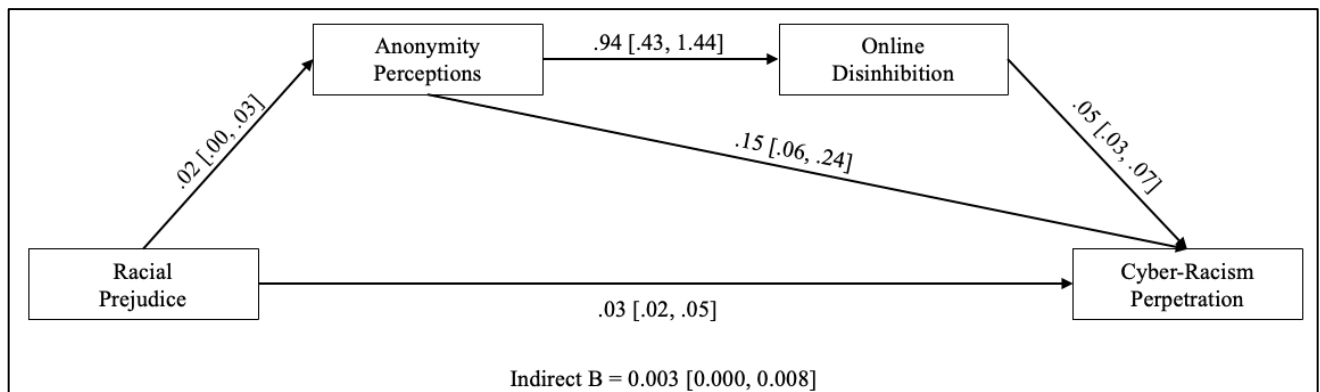


Figure 5. High Dark Personality Path Model

Though no direct hypotheses were drawn for Machiavellianism, due to the exploratory nature of the impact of this Dark Core trait, a moderated serial mediation path analysis was run to provide exploratory results in which Machiavellianism was the moderator. Results should be interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 12.52$ (2), $p = .002$, CFI = .95, TLI = .60, RMSEA = .15 [.07, .23], SRMR = .03. Machiavellianism moderated one direct path. Anonymity perceptions were only significantly associated with cyber-racism for those high, $B = 0.15$ [0.07, 0.24] but not low, $B = 0.03$ [-0.01, 0.08] in Machiavellianism. Additionally, Machiavellianism moderated the indirect serial effect, such that

the overall serial mediation only worked for participants high in Machiavellianism (see Figure 6//7).

Figure 6

Low Machiavellianism Path Model

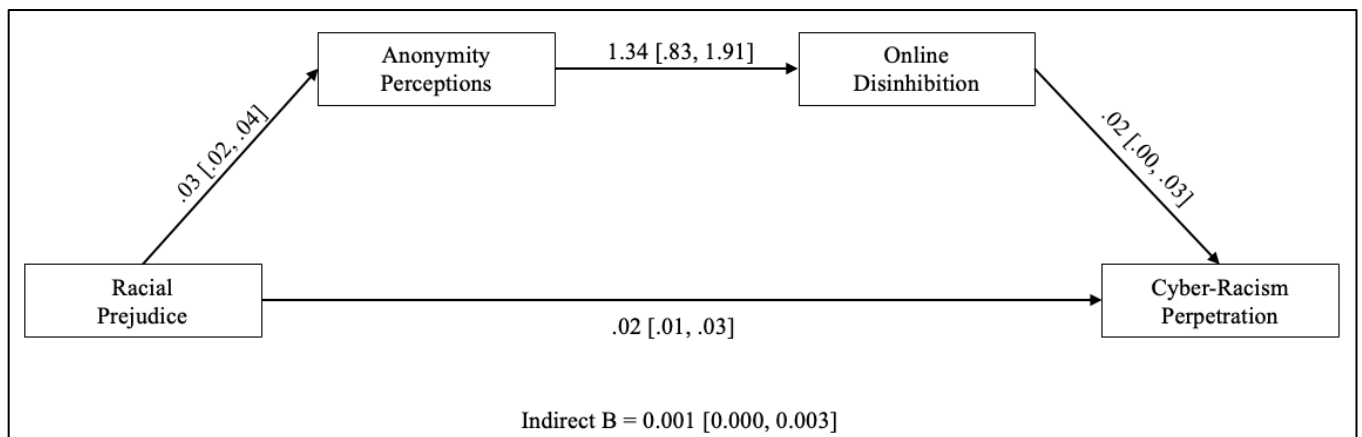


Figure 6. Low Machiavellianism Path Model

Figure 7

High Machiavellianism Path Model

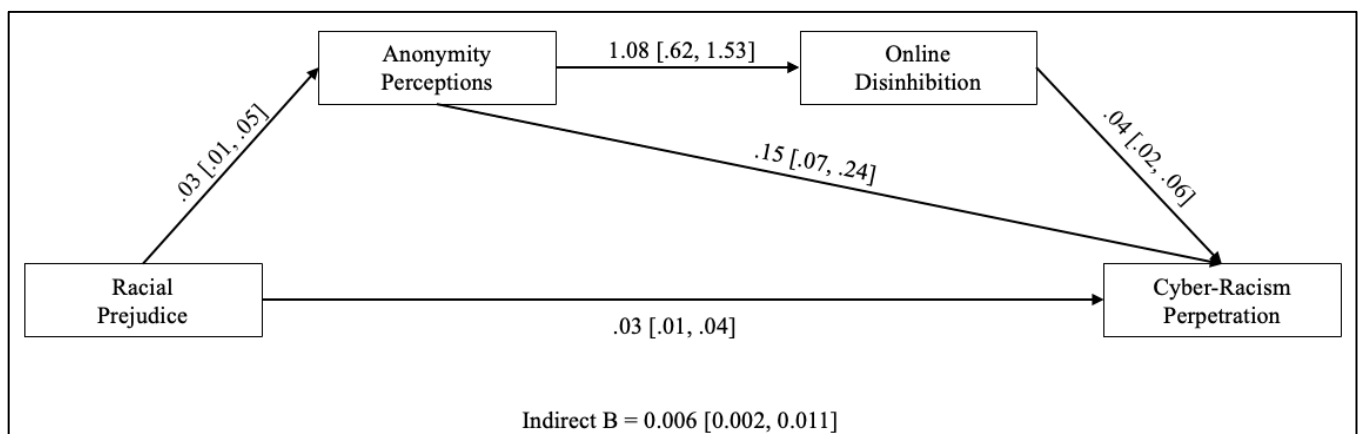


Figure 7. High Machiavellianism Path Model

Similarly, though no direct hypotheses were drawn for Narcissism, due to the exploratory nature of this variable, a moderated serial mediation path analysis was run to provide exploratory results, with Narcissism as the moderator. Results should be interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 15.99 (2)$, $p < .001$, CFI = .94, TLI = .46,

RMSEA= .17 [.09, .26], SRMR = .03. Narcissism moderated two direct paths in this model. Racial prejudice was only significantly associated with cyber-racism for those high, $B = 0.03$ [0.02, 0.04] but not low, $B = 0.01$ [0.00, 0.02] in Narcissism. Additionally, anonymity perceptions were only significantly associated with cyber-racism for those high, $B = 0.15$ [0.06, 0.25] but not low, $B = 0.03$ [-0.01, 0.07] in Narcissism. Finally, Narcissism did moderate the indirect serial effect, such that the overall serial mediation only worked for participants high in Narcissism (see Figure 8/9).

Figure 8

Low Narcissism Path Model

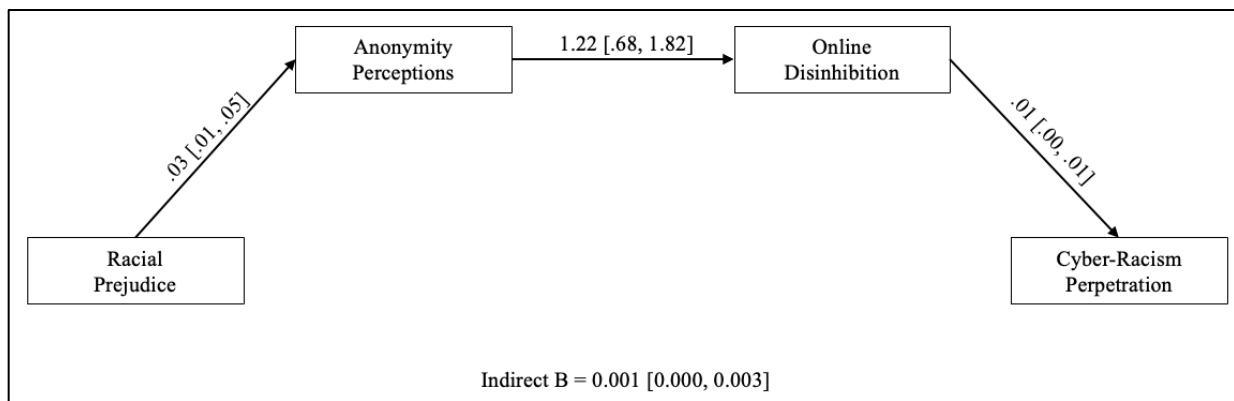


Figure 8. Low Narcissism Path Model

Figure 9

High Narcissism Path Model

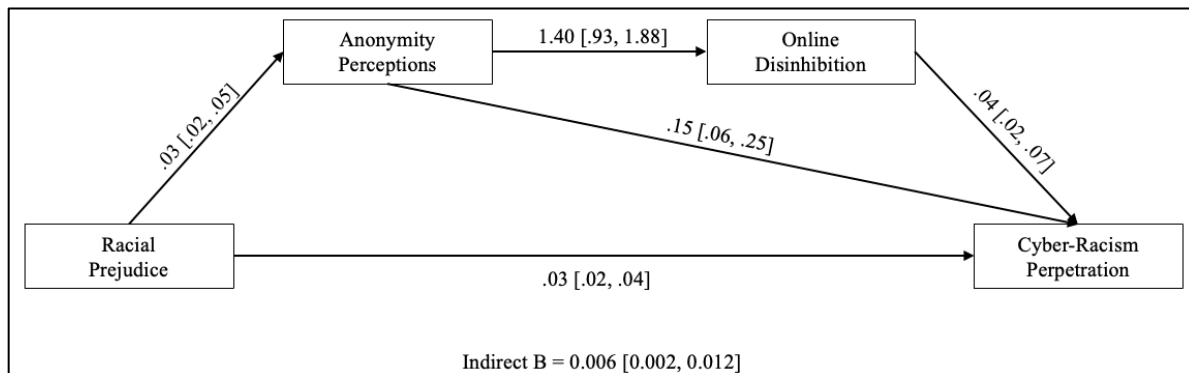


Figure 9. High Narcissism Path Model

Regarding H8, in which Psychopathy was the moderator applied to the integrated cyber-racism model, a moderated serial mediation path analysis was conducted. Results should be interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 13.08$ (2), $p = .001$, CFI = .95, TLI = .54, RMSEA = .15 [.07, .24], SRMR = .04. Psychopathy moderated several direct paths. Anonymity perceptions were significantly associated with cyber-racism only for those high, $B = 0.15$ [0.06, 0.24] but not low, $B = 0.04$ [-0.00, 0.09] in Psychopathy. Additionally, online disinhibition was significantly associated with cyber-racism for those high, $B = 0.05$ [0.03, 0.07] but not low, $B = 0.004$ [-0.00, 0.01] in Psychopathy. These results support H8 as Psychopathy moderated the direct path between anonymity perceptions and cyber-racism, as well as the indirect serial effect (see Figure 10/11). In other words, the integrated cyber-racism model worked for those high in Psychopathy but not low.

Figure 10

Low Psychopathy Path Model

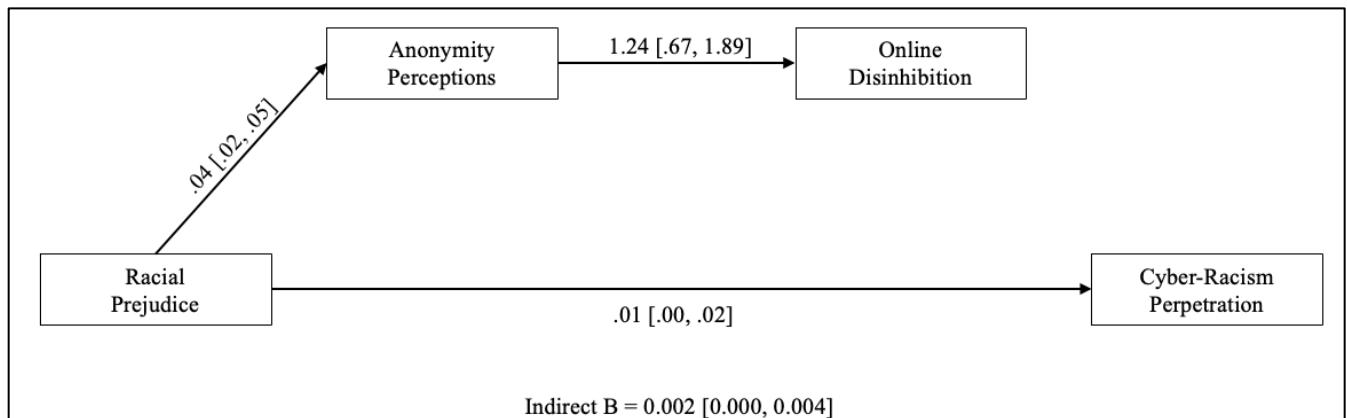


Figure 10. Low Psychopathy Path Model

Figure 11

High Psychopathy Path Model

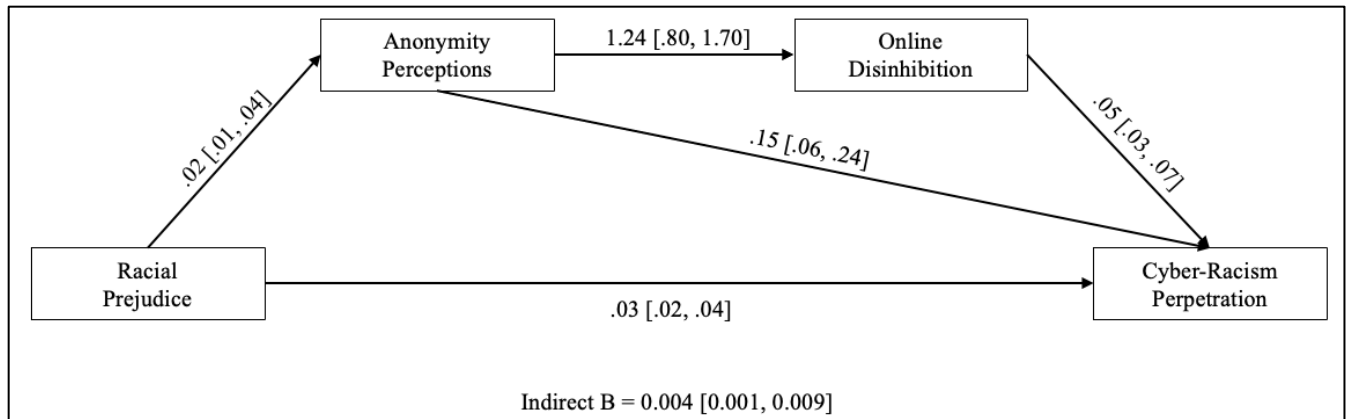


Figure 11. High Psychopathy Path Model

Finally, per H9 in which Sadism was the moderator applied to the integrated cyber-racism model, a serial mediation path analysis was conducted. Results should be interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 8.54$ (2), $p = .014$, CFI = .96, TLI = .60, RMSEA = .11 [.03, .21], SRMR = .03. Sadism moderated many direct paths. Racial prejudice was significantly associated with anonymity perceptions for those high, $B = 0.03$ [0.01, 0.04] but not low, $B = 0.02$ [0.00, 0.03] in Sadism. Racial prejudice was only significantly associated with cyber-racism for those high, $B = 0.03$ [0.02, 0.05] but not low, $B = 0.003$ [0.00, 0.01] in Sadism. Anonymity perceptions were significantly associated with cyber-racism only for those high, $B = 0.15$ [0.06, 0.26] but not low, $B = 0.01$ [-0.01, 0.04] in Sadism. Lastly, online disinhibition was significantly associated with cyber-racism for those high, $B = 0.05$ [0.03, 0.08] but not low, $B = 0.01$ [0.00, 0.01] in Sadism. These results support H9 as Sadism moderated the direct path between anonymity perceptions and cyber-racism. and the indirect serial effect, such that the integrated cyber-racism model worked for those high in Sadism but not low (see Figure 12/13).

Figure 12

Low Sadism Path Model

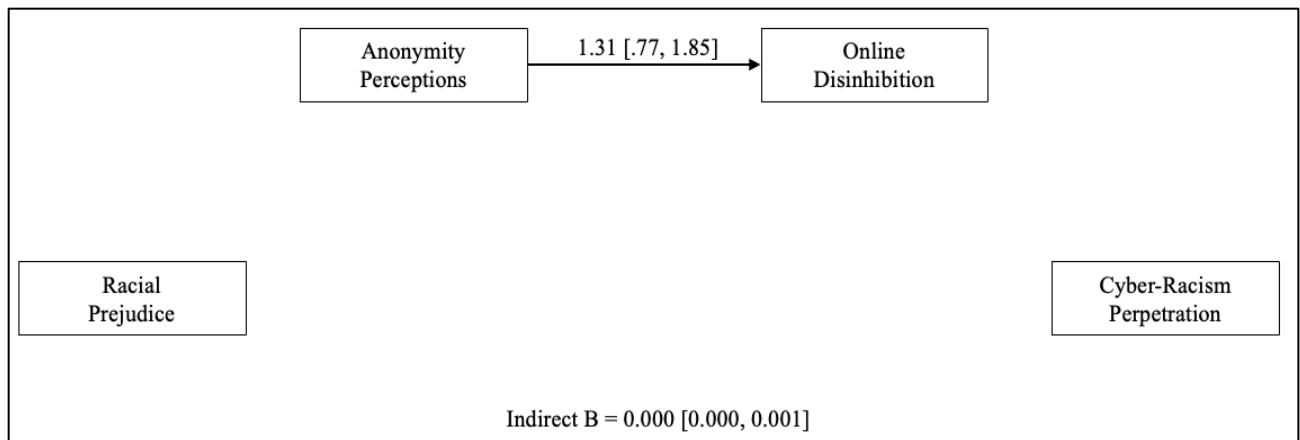


Figure 12. Low Sadism Path Model

Figure 13

High Sadism Path Model

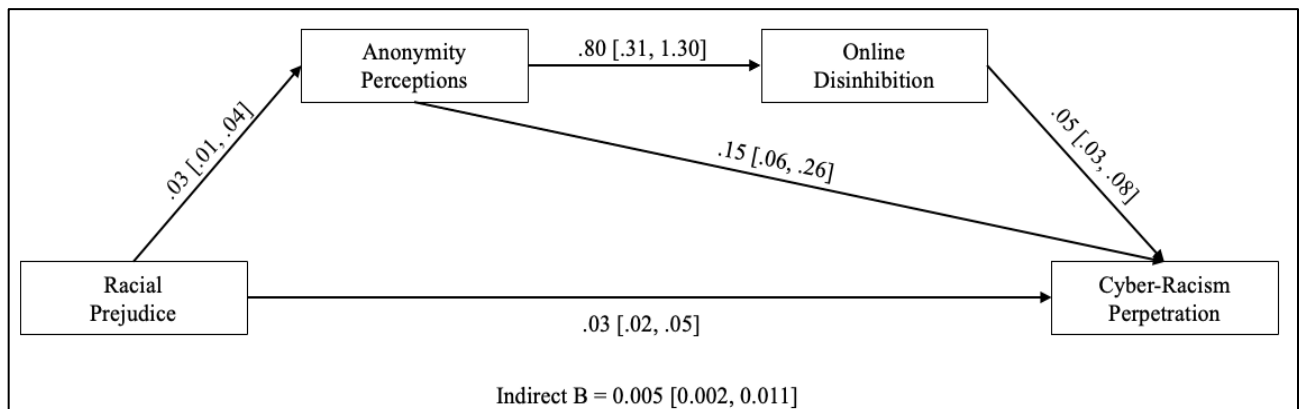


Figure 13. High Sadism Path Model

Study 1 Summary

Study 1 aimed to expand the current cyber-racism literature and theorizing by examining how individual differences in personality (i.e., the Dark Tetrad) moderates the mediating pathways in the integrated cyber-racism theory. The results found support for the base Keum and Miller (2018) cyber-racism model but more importantly, furthered our theoretical understanding of the integrated cyber-racism process by providing support for the JSM explanation of cyber-racism. Results suggest that anonymity and online disinhibition serially mediate the relationship

between racial prejudice and cyber-racism, acting as a release of the usual suppression.

Additionally, this integrated cyber-racism model was only significant predominantly for those who were high in the Dark Tetrad traits. All together, these results provide further understanding to the processes germane to cyber-racism perpetration and identify for whom this model may be most important for. To further our understanding into specific situations that make cyber-racism more likely, Study 2 builds off these findings and attempts to identify which specific dimensions of online disinhibition are the strongest mediators of the integrated cyber-racism model.

Chapter 3 - Study 2: Dimensions of Online Disinhibition

Suler's (2004; 2005) Online Disinhibition Effect operationalizes online disinhibition as a multi-dimensional construct that is comprised of six distinct dimensions that explain why people act differently online than in face-to-face interactions. Recall, these dimensions are dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, and minimization of authority. Research has predominantly focused on the impact of anonymity and invisibility cyber aggression (Lapidot-Lefler & Barak, 2012). To my knowledge, there is no published data that has tested the relationship between the six dimensions of online disinhibition and nefarious online behaviors, including cyber-aggression and cyber-racism.

Consistent with Study 1, Study 2 investigated how the Dark Tetrad impacts cyber-racism via the integration of JSM and the Keum and Miller (2018) model in order to show replication, and, therefore, we will examine the moderating role of the Dark Tetrad in the serially mediated effect of anonymity and online disinhibition in the relationship between racial prejudice and cyber-racism. To do this, we will first operationalize online disinhibition as the composite of all six dimensions articulated by Suler (2004, 2005), and I hypothesize that the results from Study 1 will replicate. Then, Study 2 will build upon the findings of Study 1 by simultaneously analyzing

the six dimensions of online disinhibition in a single model to add further clarification regarding what specific online disinhibition dimensions are most important for predicting cyber-racism within the context of our theoretically integrated cyber-racism model found in Study 1. To do this, participants will complete measures of the different dimensions of online disinhibition, the Dark Tetrad, racial prejudice, anonymity, and cyber-racism perpetration. We are unaware of any published study examining these effects.

Study 2 Hypotheses

The hypotheses for Study 2 pertain to the general relationships between the Dark Tetrad, racial prejudice, anonymity perceptions, the dimensions of online disinhibition, and cyber-racism perpetration. The first set of hypotheses will outline the anticipated correlations between the variables I measured.

H1: The Dark Tetrad traits will be positively correlated with racial prejudice, anonymity perceptions, each online disinhibition dimension, and cyber-racism.

H2: Racial prejudice will be positively correlated with anonymity, each online disinhibition dimension, and cyber-racism.

H3: Anonymity will be positively correlated with each online disinhibition dimension and cyber-racism.

H4: All online disinhibition dimensions will be positively correlated with cyber-racism.

Next, the subsequent hypotheses focus on our prediction regarding the mediating effects inherent in the juxtaposition of JSM (Crandall & Eshleman, 2003), the Keum and Miller (2018) model, and specific dimensions of online disinhibition. Please note that Figure 9 depicts the conceptual model for ease of readability.

H5: The dissociative anonymity dimension of online disinhibition will positively partially mediate the relationship between anonymity and cyber-racism.

H6: The asynchrony dimension of online disinhibition will positively partially mediate the relationship between anonymity and cyber-racism.

H7: The minimization of authority dimension of online disinhibition will positively partially mediate the relationship between anonymity and cyber-racism.

H8: Anonymity and the dissociative anonymity dimension of online disinhibition will positively partially serially mediate the relationship between racial prejudice and cyber-racism.

H9: Anonymity and the asynchrony dimension of online disinhibition will positively partially serially mediate the relationship between racial prejudice and cyber-racism.

H10: Anonymity and the minimization of authority dimension of online disinhibition will positively partially serially mediate the relationship between racial prejudice and cyber-racism.

Finally, the next hypotheses will predict the moderating role of the Dark Tetrad on the hypothesized serially mediated pathways from genuine prejudice to cyber-racism through anonymity perceptions and various dimensions of online disinhibition.

H11: The Dark Tetrad (as a composite and portioned into its subscales) will moderate the mediating pathways of dissociative anonymity, asynchrony, and minimization of authority, and the direct pathway of anonymity and cyber-racism such that this model will be especially strong for participants high on the Dark Personality compared to participants who are low.

H12: Psychopathy will moderate the mediating pathways of dissociative anonymity, asynchrony, and minimization of authority, and the direct pathway of anonymity and cyber-racism such that this model will be especially strong for participants high on psychopathy compared to participants who are low.

H13: Sadism will moderate the mediating pathways of the dissociative anonymity, asynchrony, and minimization of authority, and the direct pathway of anonymity and cyber-racism such that this model is especially strong for participants high on sadism compared to participants who are low.

These hypotheses stem from the theorized Keum and Miller (2018) model, the JSM (Crandall & Eshleman, 2003) and Suler's (2004; 2005) distinctions. The theoretical ordering for the model stems from the integrated cyber-racism theory found in Study 1. No published study has investigated all dimensions of online disinhibition in their ability to predict and mediate the relationships around cyber-racism.

Figure 14
Study 2 Conceptual Model

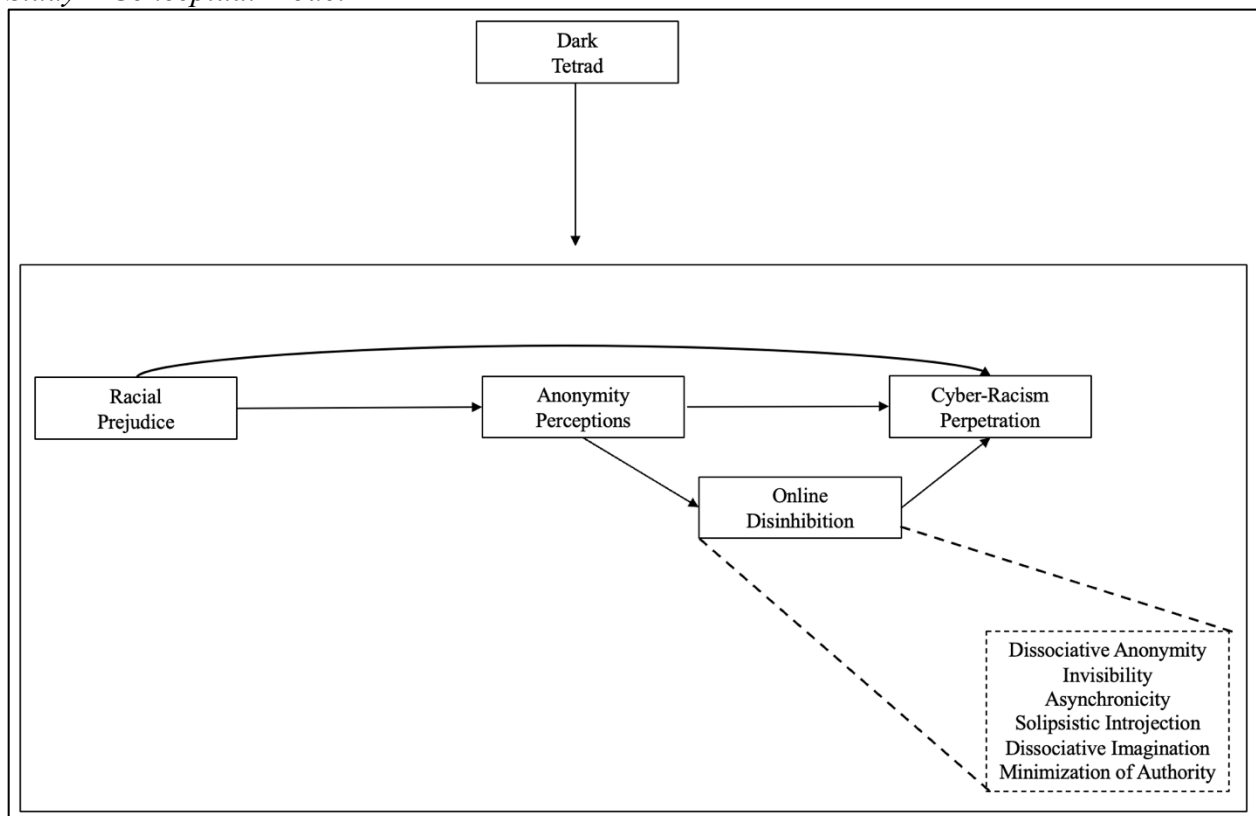


Figure 14. Study 2 Conceptual Model

Method

Participants and Procedure

According to an a priori power analysis (semPower; Moshagen & Bader, 2024), 250 participants were needed to achieve .80 power using the AGFI = .90, when the model has 10 variables ($df = 6$). To ensure sufficient data quality, attention checks were included in the survey. Failure of the attention checks resulted in the participant being dropped from analysis. To account for these potential drops, 609 participants were recruited from CloudResearch in exchange for \$2.00 (US) each. The inclusionary criteria required participants to be 18 years or older and be in the US.

Of the 609 participants who completed the survey, several were omitted from analysis due to failure of the reCAPTCHA element ($n = 13$), failure to complete multiple survey items ($n = 4$), and failure of one or more attention checks ($n = 23$). These data cleaning procedures resulted in 569 participants for data analysis. On average, participants were 40.62 years old ($SD = 12.20$) and majority White (72.8%). Regarding sex, 49.9% identified as men and 48.7% as women. Therefore, we were sufficiently powered.

Participants completed a series of surveys assessing several constructs, including demographic information, the Dark Tetrad, anonymity perception, online disinhibition, social desirability, racial prejudice, and cyber-racism perpetration using Qualtrics, mostly identical to Study 1. The key differences between Studies 1 and 2 was how online disinhibition is operationalized, such the measure of online disinhibition in Study 2 measures the six dimensions of online disinhibition and the sum of all items yields an overall online disinhibition score. Lastly, participants were debriefed, thanked, and compensated.

Materials

The same Dark Tetrad (Mach: $\alpha = .77$, Narc: $\alpha = .84$, Psychopathy: $\alpha = .83$, Sadism: $\alpha = .82$, Total: $\alpha = .89$), anonymity perceptions ($\alpha = .91$), racial prejudice ($\alpha = .93$), social desirability ($\alpha = .81$), cyber-racism perpetration ($\alpha = .95$), and demographic questionnaires from Study 1 were used in Study 2 (See Study 1 Materials). In addition, the following questionnaire was also used.

Online Disinhibition. To measure the degree to which participants feel or behave differently in online interactions, the 23-item Online Disinhibition Scale (Cheung et al., 2020; See Appendix G) was used. This measure asked participants to rate their level of agreement on a 1 (*strongly disagree*) to 5 (*strongly agree*) rating scale for each item. During scale development (Cheung et al., 2020), this scale demonstrated psychometric properties (e.g., construct validity, reliability, internal consistency, and clear distinction of six dimensions of online disinhibition). This was my preferred measure due to its distinction of separate dimensions of online disinhibition, instead of focusing on a single specific behavior. This composite measure had acceptable reliability ($\alpha = .93$).

This online disinhibition scale identifies six subscales of online disinhibition, ranging from three to five items within each. These subscales include four dissociative anonymity items, $\alpha = .94$ (*My actions are not identifiable in the online environment*), five invisibility items, $\alpha = .95$ (*I feel invisible in the online environment*), three asynchronicity items, $\alpha = .93$ (*I can delay my feedback to others in the online environment*), three solipsistic introjection items, $\alpha = .68$ (*I assign a character to that person I am communicating with online*), four dissociative imagination items, $\alpha = .87$ (*The online environment has no connection to reality*), and four minimization of authority items, $\alpha = .92$ (*I feel less fear of offline authorities in the online environment*). For

scoring participants' responses, I created composites for each of the subscales, given that these were designed to identify specific dimensions of online disinhibition (Cheung et al., 2020).

Data Analysis Plan

All variables were centered before inclusion in the model, and one covariate was included: social desirability. For Study 2, correlations were calculated using a Spearman correlation, as the data was non-normal. The mediation and subsequent moderated mediation analyses were conducted using the lavaan package in R (Rosseel, 2012). For the moderated mediation analyses, a multi-group path analysis was conducted using a median split⁶ (i.e., High vs. Low) of the Dark Tetrad moderators. A MLR/Robust correction (Rosseel, 2012) was used to estimate model fit as the data was non-normal and bootstrapping was used to normalize the data.

Results

Correlations

Table 2 displays the zero-order correlations and descriptive statistics between the continuous variables (e.g., the Dark Tetrad, racial prejudice, anonymity perceptions, online disinhibition dimensions, and cyber-racism). In line with H1, all Dark Tetrad variables were (significantly) positively correlated with racial prejudice, anonymity, most dimensions of online disinhibition, and cyber-racism. H2 was supported for the most part, as racial prejudice was positively correlated with anonymity, most dimension of online disinhibition, and cyber-racism. Similarly, H3 and H4 were supported such that, both anonymity and most of the dimensions of online disinhibition were positively correlated with each other and with cyber-racism. The lone

⁶ I acknowledge that median splits hold several large limitations, but a multi-group path analysis approach was used for parsimony and to extract meaningful results from complex models. Prior to a median split approach, attempts were made to split the data by +/- 1SD but due to a floor effect and no variance in cyber-racism perpetration, the model would not converge.

exception to these hypotheses was the asynchronicity dimension of online disinhibition.

Asynchronicity was not correlated with racial prejudice, cyber-racism, narcissism, psychopathy, and two other dimensions of online disinhibition (e.g., minimization of authority and dissociative imagination).

Table 2*Study 2 Correlations and Descriptive Statistics*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1: Anonymity Perceptions	----													
2: Dissociative Anonymity (OD)	.55**	----												
3: Invisibility (OD)	.49**	.82**	----											
4: Asynchronicity (OD)	.12**	.19**	.13**	----										
5: Solipsistic Introjection (OD)	.30**	.37**	.33**	.19**	----									
6: Dissociative Imagination (OD)	.36**	.42**	.48**	-.05	.28**	----								
7: Authority Minimization (OD)	.56**	.58**	.59**	.02	.32**	.54**	----							
8: Cyber-Racism Perpetration	.29**	.22**	.20**	-.06	.18**	.27**	.28**	----						
9: Racial Prejudice	.32**	.23**	.26**	-.11*	.19**	.42**	.32**	.32**	----					
10: Machiavellianism	.29**	.24**	.26**	.11*	.29**	.28**	.29**	.20**	.22**	----				
11: Narcissism	.24**	.18**	.15**	-.02	.17**	.21**	.20**	.25**	.17**	.29**	----			
12: Psychopathy	.27**	.16**	.18**	.03	.11**	.27**	.24**	.32**	.22**	.24**	.42**	----		
13: Sadism	.41**	.29**	.28**	.21**	.24**	.26**	.34**	.36**	.29**	.36**	.24**	.56**	----	
14: Social Desirability	-.05	.02	.02	-.17**	-.06	.05	-.02	-.05	.06	-.06	.09*	-.14**	-.16**	----
<i>M</i>	6.71	10.01	11.55	11.66	8.83	7.97	8.43	5.98	53.74	23.22	18.76	13.46	16.68	8.83
<i>SD</i>	3.21	4.69	5.64	3.20	2.54	3.84	4.18	3.11	28.00	4.68	5.66	5.52	6.19	3.80

Note: ** $p < .01$, * $p < .05$

Table 2. Study 2 Correlations

Mediation

The basic Keum and Miller (2018) model was tested again to establish further replication. In this analysis, anonymity was the independent variable, the total online disinhibition score was the mediator, and cyber-racism was the dependent variable. It should be noted that all paths were estimated, which resulted in no degrees of freedom, and, therefore, PROCESS (Hayes, 2022) was used to estimate the model instead of the *laavan* package in R. Results supported the Keum and Miller (2018) model and replicated Study 1 results. In terms of the individual paths, anonymity perceptions were significantly associated with the total online disinhibition ($B = 3.32$, $SE = 0.18$, $p < .001$). Additionally, both anonymity perceptions ($B = 0.23$, $SE = 0.04$, $p < .001$) and total online disinhibition ($B = 0.04$, $SE = 0.01$, $p < .001$) were significantly associated with cyber-racism. In line with the results of Study 1, the indirect effect of the total online disinhibition was significant (*Indirect* $B = 0.12$ [.07, .18]). Anonymity perceptions significantly predicted online disinhibition to predict subsequent cyber-racism perpetration.

A serial mediation path analysis was additionally run in PROCESS (Hayes, 2022) with the total online disinhibition score as one of the mediators in the analysis to show replication from Study 1. In terms of the individual paths, racial prejudice was significantly associated with anonymity perceptions ($B = 0.04$, $SE = 0.005$, $p < .001$). Anonymity perceptions were significantly associated with online disinhibition ($B = 3.07$, $SE = 0.19$, $p < .001$). Additionally, racial prejudice ($B = 0.03$, $SE = 0.004$, $p < .001$), anonymity perceptions ($B = 0.17$, $SE = 0.01$, $p < .001$) and online disinhibition ($B = 0.02$, $SE = 0.01$, $p = .021$) were significantly associated with cyber-racism. In line with the Study 1 results, the serial indirect effect of anonymity perceptions and online disinhibition was significant (*Indirect* $B = 0.002$ [.001, .004]). In other

words, racial prejudice predicted anonymity which in turn predicted online disinhibition, which predicted subsequent cyber-racism perpetration.

To investigate further into which specific dimensions of online disinhibition mediate the relationship between anonymity and cyber-racism, H5 posited that dissociative anonymity would mediate the relationship between anonymity and cyber-racism. The results did not support for H5, as the indirect effect of the dissociative anonymity online disinhibition was not significant (*Indirect B* = - 0.001 [-.05, .05]).

Regarding H6, which posited that asynchrony would mediate the relationship between anonymity and cyber-racism, results were mixed. In partial support of H6, the indirect effect of the asynchrony online disinhibition was significant (*Indirect B* = -0.03 [-.03, -.01]) but showed a negative relationship rather than the hypothesized positive relationship.

Finally in regard to H7, which posited that minimization of authority would mediate the relationship between anonymity and cyber-racism was supported. The indirect effect of the minimization of authority online disinhibition was significant (*Indirect B* = 0.10 [.01, .11]). Anonymity predicted minimization of authority which predicted subsequent cyber-racism perpetration.

Serial Mediation

To investigate the integrated cyber-racism theory and which dimensions of online disinhibition acted within this model, H8 posited that anonymity and dissociative anonymity would partially serially mediate the suggested model. Results found that H8 was not supported as the serial indirect effect of anonymity perceptions and dissociative anonymity online disinhibition was not significant (*Indirect B* = 0.000 [-.002, .002]).

Regarding H9, which posited that anonymity and asynchrony would partially serially mediate the relationship between racial prejudice and cyber-racism, had mixed results. Results suggest that H9 was not supported as the serial indirect effect of anonymity perceptions and asynchrony online disinhibition was not significant (*Indirect B* = -0.001 [-.001, .000]).

Finally in regard to H10, in which anonymity and minimization of authority would mediate the relationship between racial prejudice and cyber-racism, results were not significant. H10 was not supported as the serial indirect effect of anonymity perceptions and minimization of authority online disinhibition was not significant (*Indirect B* = 0.002 [.000, .004]).

Moderated Mediation⁷

To test how the Dark Tetrad (combined and subscales) moderate the integrated cyber-racism model with the separate dimensions of online disinhibition, several multi-group path models were conducted. Regarding H11, which posited that the total Dark Personality score would moderate the Study 2 conceptual model, a moderated serial mediation path analysis was run. Results should be interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 54.05$ (12), $p < .001$, CFI = .97, TLI = .77, RMSEA = .12 [.09, .16], SRMR = .05. The Total Dark Personality moderated five direct paths. Anonymity perceptions were significantly associated with the asynchrony dimension of online disinhibition for those low, $B = 0.24$ [0.10, 0.37] but not high, $B = 0.08$ [-0.04, 0.19] in Total Dark Personality. Racial prejudice was significantly associated with cyber-racism for those high, $B = 0.04$ [0.02, 0.06] but not low, $B = 0.002$ [0.00, 0.01] in Total Dark Personality. Anonymity perceptions were significantly associated with cyber-racism for those high, $B = 0.16$ [0.06, 0.28] but not low, $B = 0.03$ [-0.00, 0.09] in Total Dark Personality. Furthermore, regarding the dimensions of online disinhibition,

⁷ All six dimensions were included in the analyses. For ease of readability, only the hypothesized and/or significant relationships were depicted in the following figures.

both asynchrony and solipsistic introjection were only significantly associated with cyber-racism for those high, $B = -0.20 [-0.34, -0.08]$; $B = 0.14 [0.02, 0.27]$ but not low, $B = -0.02 [-0.07, 0.01]$; $B = 0.003 [-0.02, 0.03]$ in Total Dark Personality. These results partially support H11 as Total Dark Personality moderated the direct path between anonymity perceptions and cyber-racism. However, Total Dark Personality did not moderate any of the indirect serial effects which did not support H11 (see Figure 15/16).

Figure 15

Low Dark Personality Path Model

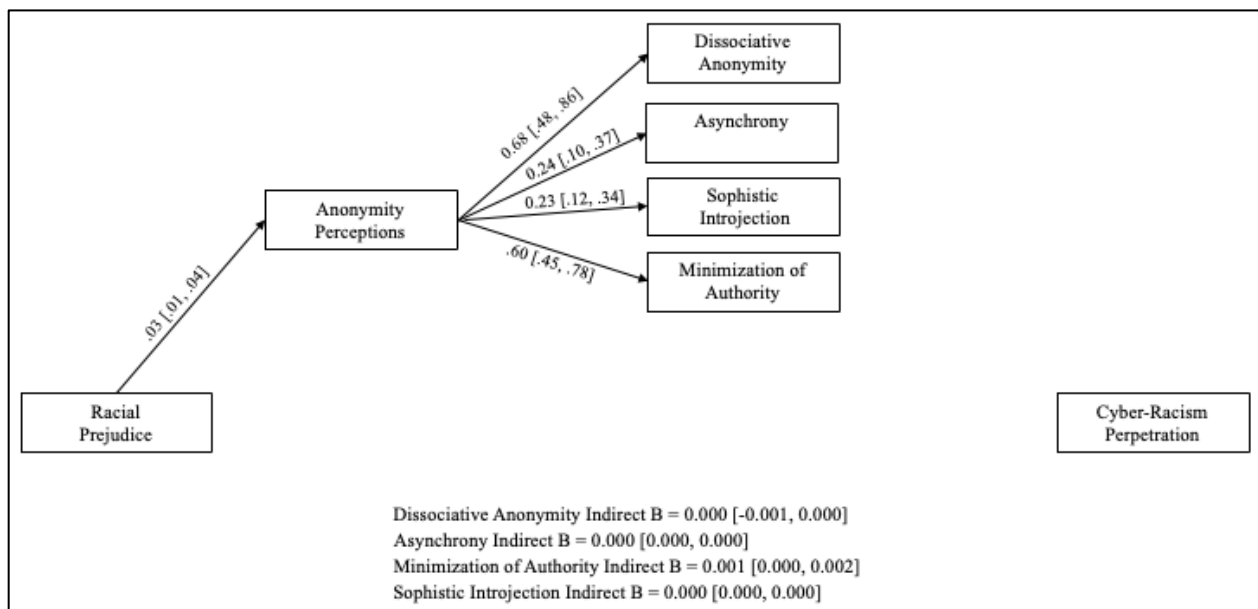


Figure 15. Low Dark Personality Path Model

Figure 16

High Dark Personality Path Model

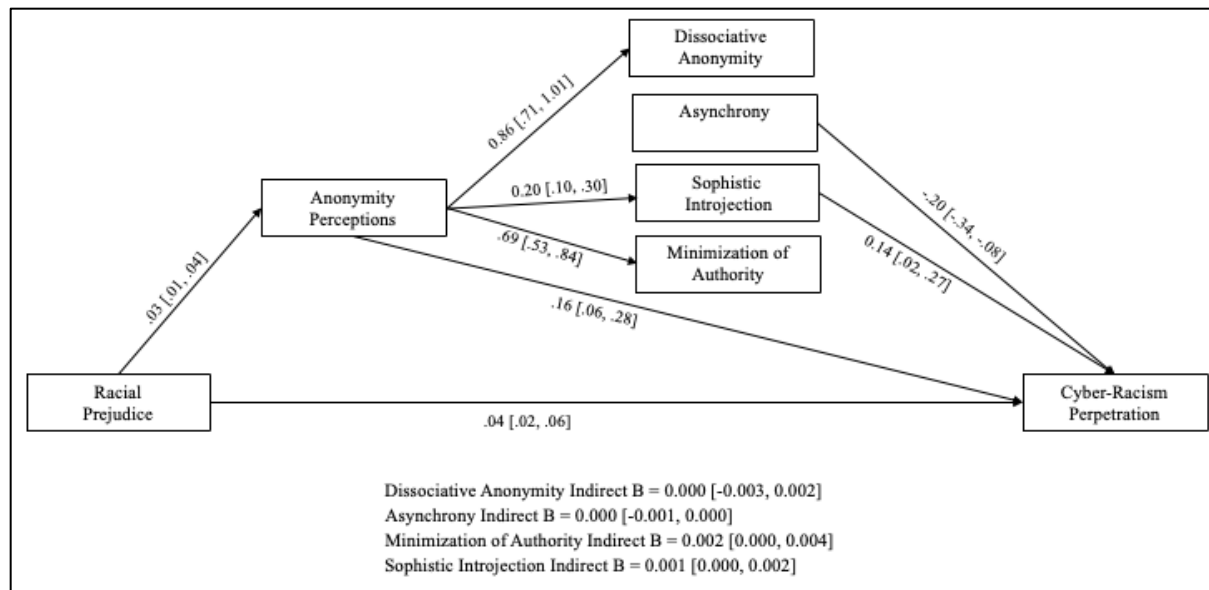


Figure 16. High Dark Personality Path Model

Though no direct hypotheses were drawn in which Narcissism was the moderator, a moderated serial mediation path analysis was run to provide exploratory results. Results should be interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 62.77 (12)$, $p < .000$, CFI = .96, TLI = .73, RMSEA = .14 [.11, .17], SRMR = .05. Narcissism moderated five direct paths. Anonymity perceptions were significantly associated with the asynchrony dimension of online disinhibition for those low, $B = 0.29 [0.18, 0.39]$ but not high, $B = 0.11 [-0.01, 0.22]$ in Narcissism. Racial prejudice was significantly associated with cyber-racism for those high, $B = 0.03 [0.02, 0.05]$ but not low, $B = 0.01 [-0.00, 0.02]$ in Narcissism. Additionally, anonymity perceptions were only significantly associated with cyber-racism for those high, $B = 0.12 [0.01, 0.23]$ but not low, $B = 0.06 [0.00, 0.15]$ in Narcissism. Furthermore, regarding the dimensions of online disinhibition, both asynchrony and minimization of authority were only significantly associated with cyber-racism for those high, $B = -0.17 [-0.29, -0.06]$; $B = 0.17 [0.07, 0.27]$ but not low, $B = -0.01 [-0.06, 0.03]$; $B = -0.05 [-0.12, -0.01]$ in Narcissism. Lastly, Narcissism moderated the indirect serial effect through the dimension of minimization of authority (see Figure 17/18).

Figure 17

Low Narcissism Path Model

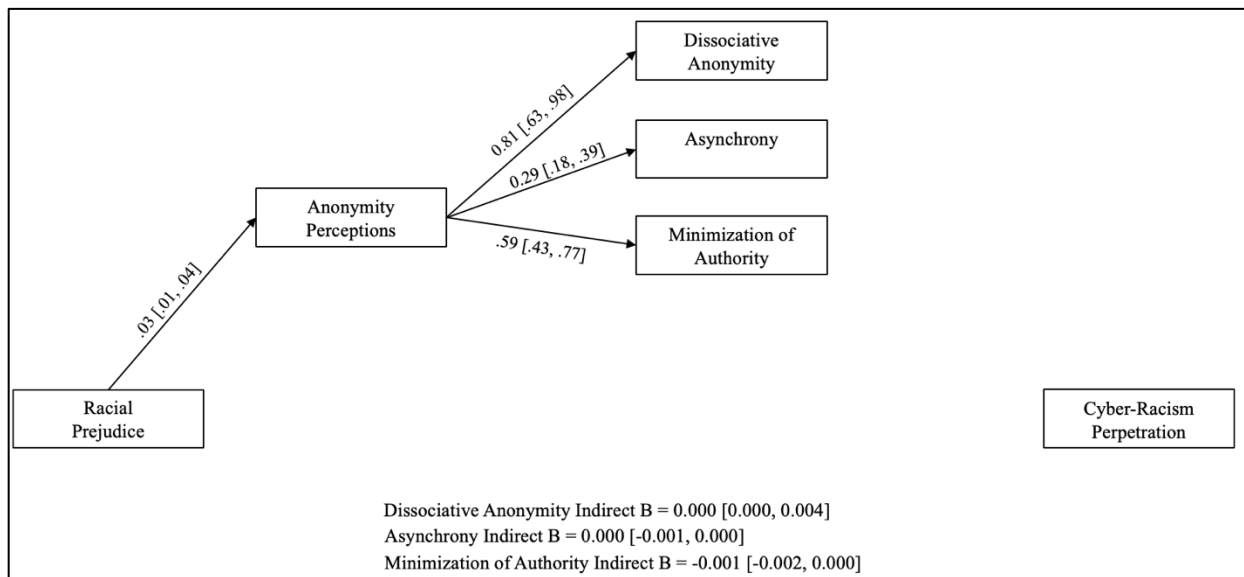


Figure 17. Low Narcissism Path Model

Figure 18

High Narcissism Path Model

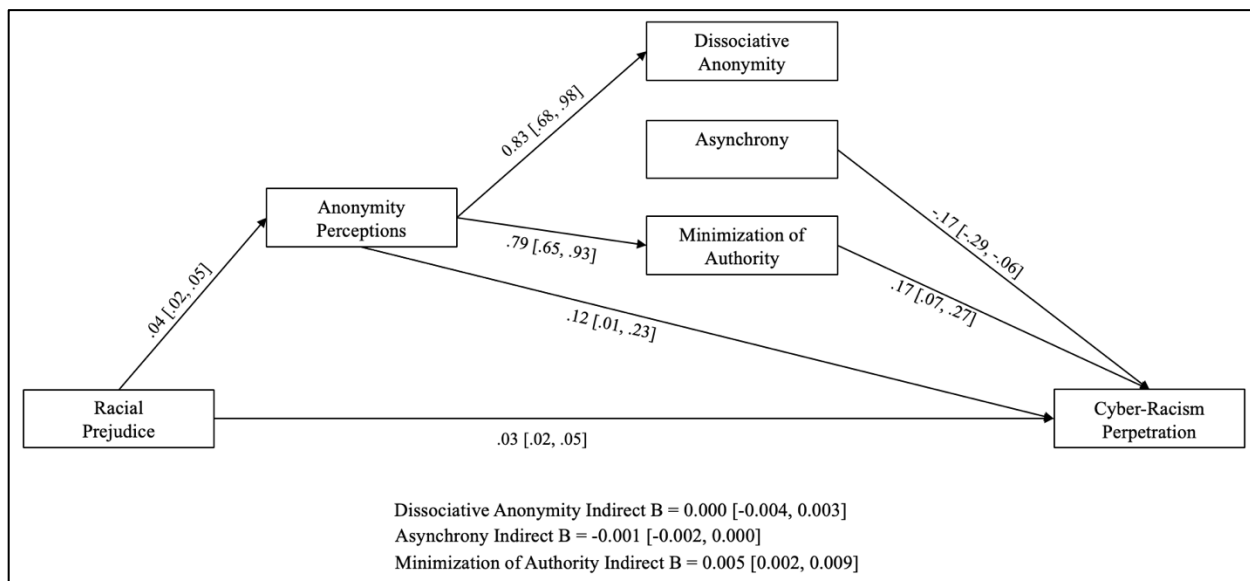


Figure 18. High Narcissism Path Model

Similarly, though no direct hypotheses were drawn for Machiavellianism as the moderator of the integrated cyber-racism model, a moderated serial mediation path analysis was run to provide exploratory results. Results should be interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 59.76$ (12), $p < .000$, CFI = .96, TLI = .74, RMSEA = .13 [.10, .17], SRMR = .05. Machiavellianism moderated two direct paths. Anonymity perceptions were significantly associated with the asynchrony dimension of online disinhibition for those high, $B = 0.18$ [0.08, 0.28] and not low, $B = 0.12$ [-0.04, 0.27] in Machiavellianism. Anonymity perceptions were only significantly associated with cyber-racism for those high, $B = 0.14$ [0.05, 0.24] but not low, $B = 0.06$ [-0.01, 0.16] in Machiavellianism. Lastly, Machiavellianism did not moderate any indirect serial effects (see Figure 19/20).

Figure 19

Low Machiavellianism Path Model

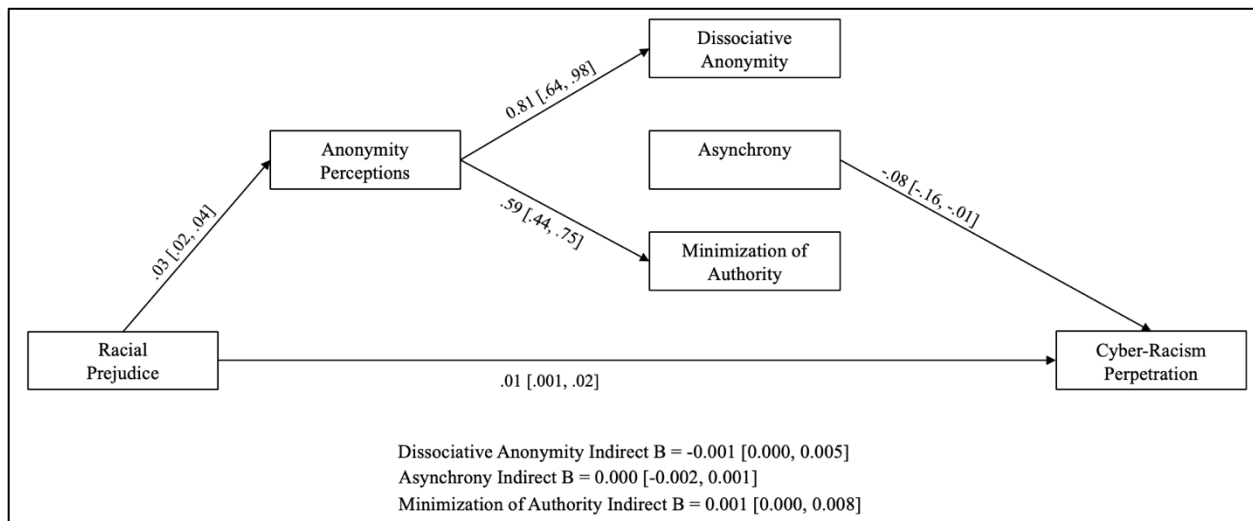


Figure 19. Low Machiavellianism Path Model

Figure 20

High Machiavellianism Path Model

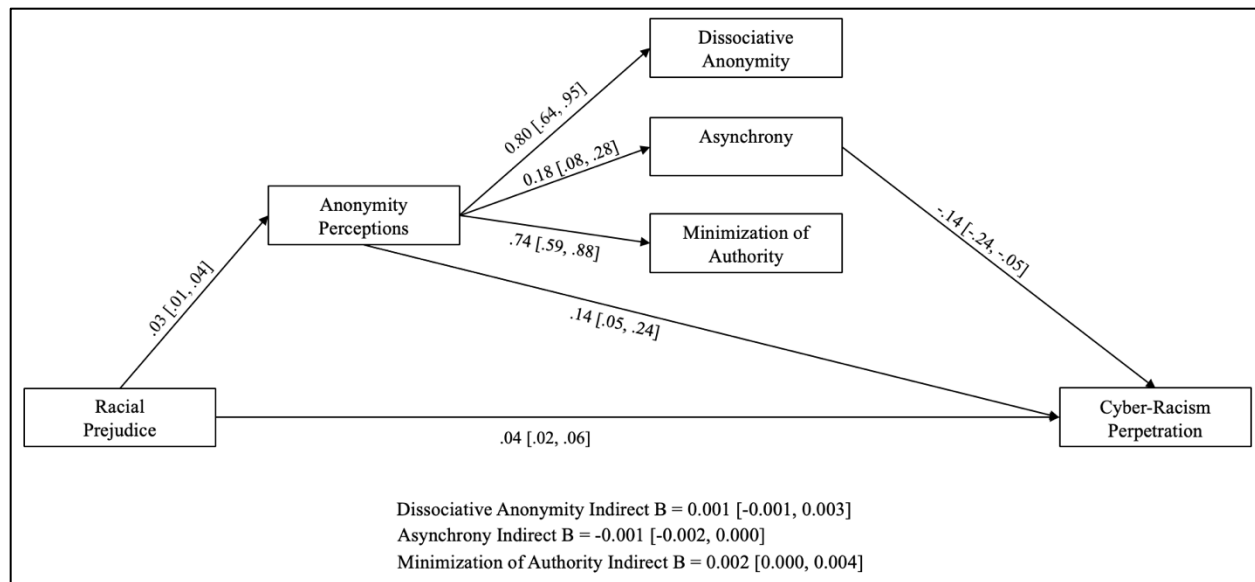


Figure 20. High Machiavellianism Path Model

Regarding H12, in which Psychopathy was the moderator applied to the Study 2 conceptual model, a moderated serial mediation path analysis was conducted. Results should be interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 56.01$ (12), $p < .001$, CFI = .97, TLI = .76, RMSEA = .13 [.09, .16], SRMR = .05. Psychopathy moderated several direct paths. Racial prejudice was significantly associated with cyber-racism for those high, $B = 0.04$ [0.02, 0.05] but not low, $B = 0.003$ [0.00, 0.01] in Psychopathy. Anonymity perceptions were significantly associated with cyber-racism for those high, $B = 0.17$ [0.06, 0.30] but not low, $B = 0.02$ [-0.01, 0.08] in Psychopathy. Furthermore, regarding the dimensions of online disinhibition, both asynchrony and solipsistic introjection were only significantly associated with cyber-racism for those high, $B = -0.18$ [-0.30, -0.07]; $B = 0.14$ [0.01, 0.29] but not low, $B = -0.02$ [-0.07, 0.02]; $B = -0.01$ [-0.03, 0.02] in Psychopathy. These results partially support H12 as Psychopathy moderated the direct path between anonymity perceptions and cyber-racism. However, Psychopathy did not moderate any of the indirect serial effects which did not support H12 (see Figure 21/22).

Figure 21

Low Psychopathy Path Model

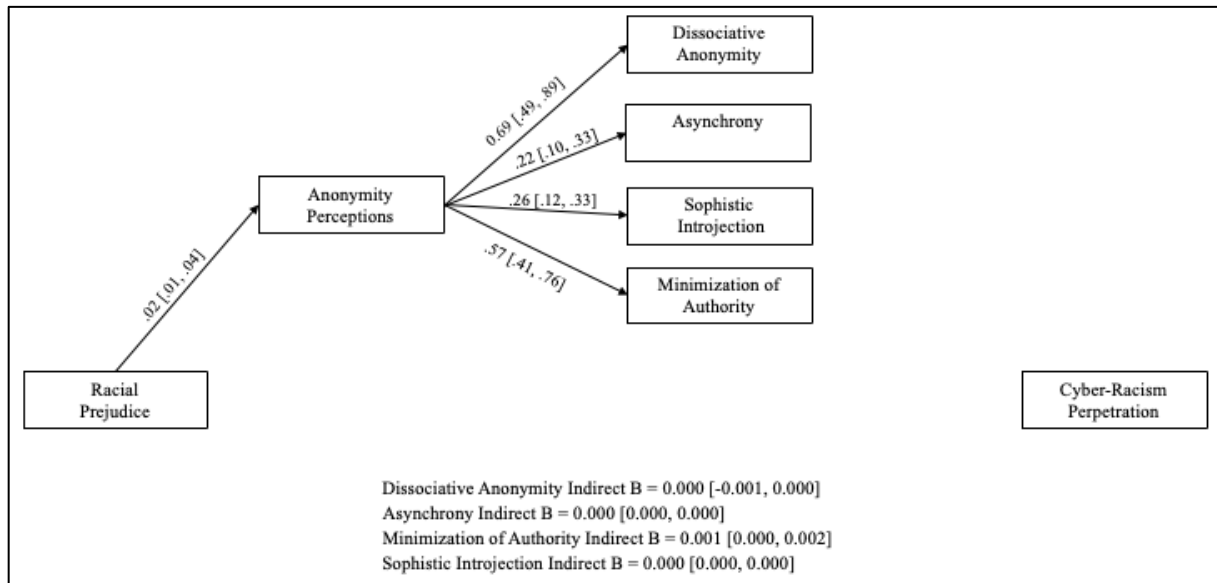


Figure 21. Low Psychopathy Path Model

Figure 22

High Psychopathy Path Model

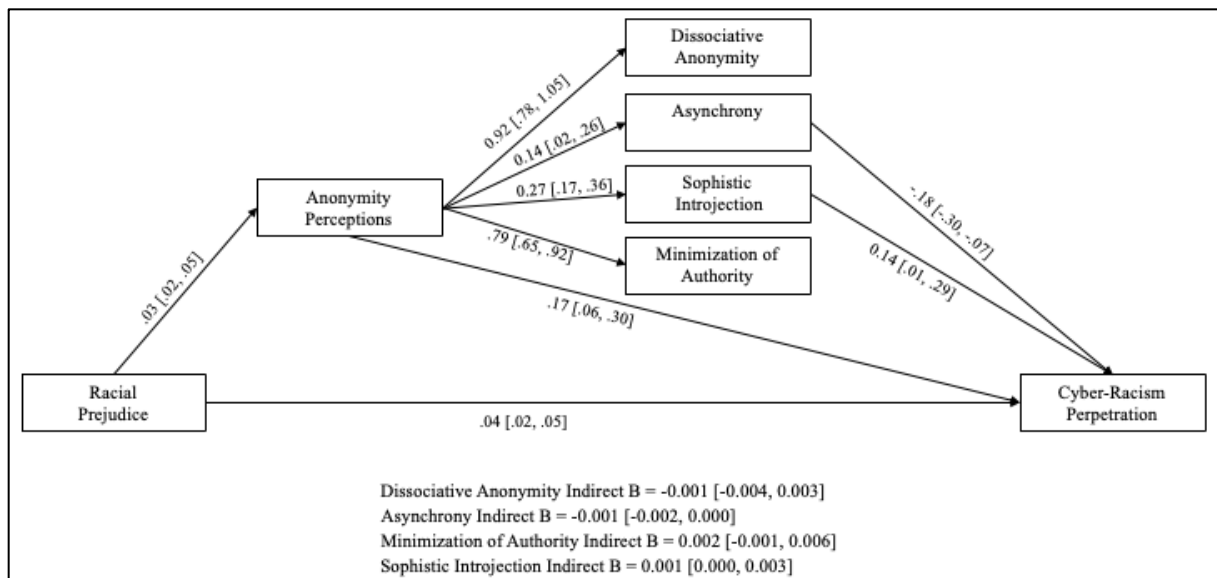


Figure 22. High Psychopathy Path Model

Finally, in terms of H13, which posited that Sadism would moderate the Study 2 conceptual model, a moderated serial mediation path analysis was conducted. Results should be

interpreted with caution due to some components of model fit not meeting thresholds, $\chi^2 = 58.27$ (12), $p < .001$, CFI = .96, TLI = .74, RMSEA = .13 [.10, .16], SRMR = .05. Sadism moderated three direct paths. Racial prejudice was significantly associated with cyber-racism for those high, $B = 0.04$ [0.02, 0.05] but not low, $B = 0.003$ [0.00, 0.01] in Sadism. Anonymity perceptions were significantly associated with cyber-racism for those high, $B = 0.17$ [0.07, 0.28] but not low, $B = 0.01$ [-0.02, 0.07] in Sadism. Furthermore, regarding the dimensions of online disinhibition, asynchrony was only significantly associated with cyber-racism for those high, $B = -0.23$ [-0.35, -0.11] but not low, $B = -0.01$ [-0.06, 0.01] in Sadism. These results partially support H13 as Sadism moderated the direct path between anonymity perceptions and cyber-racism. However, Sadism did not moderate any of the indirect serial effects which did not support H13 (see Figure 23/24).

Figure 23

Low Sadism Path Model

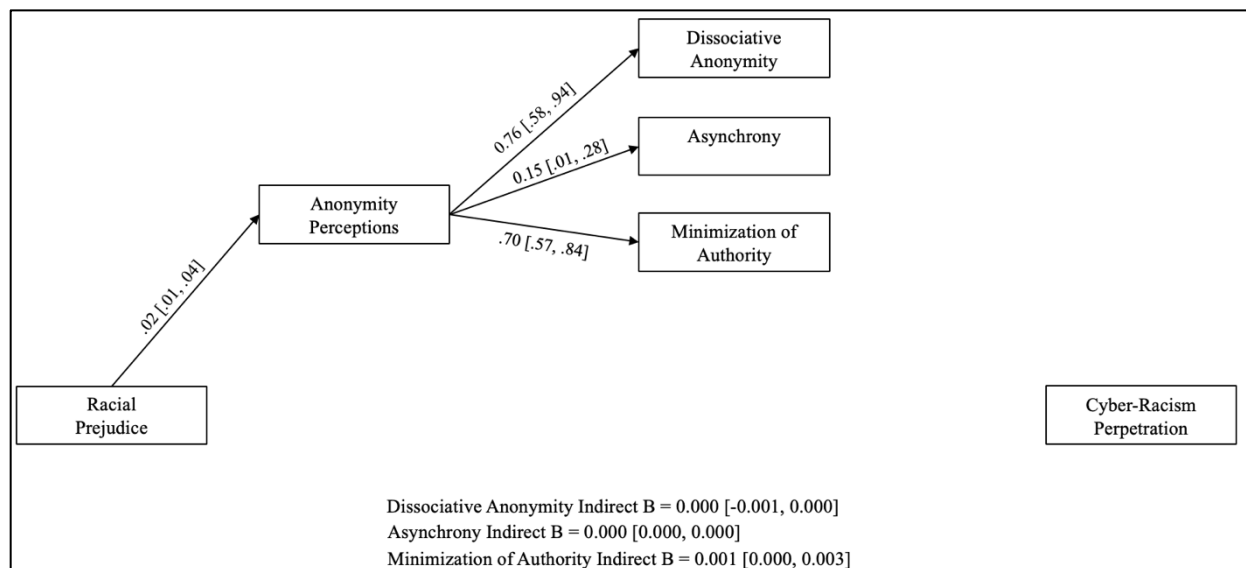


Figure 23. Low Sadism Path Model

Figure 24

High Sadism Path Model

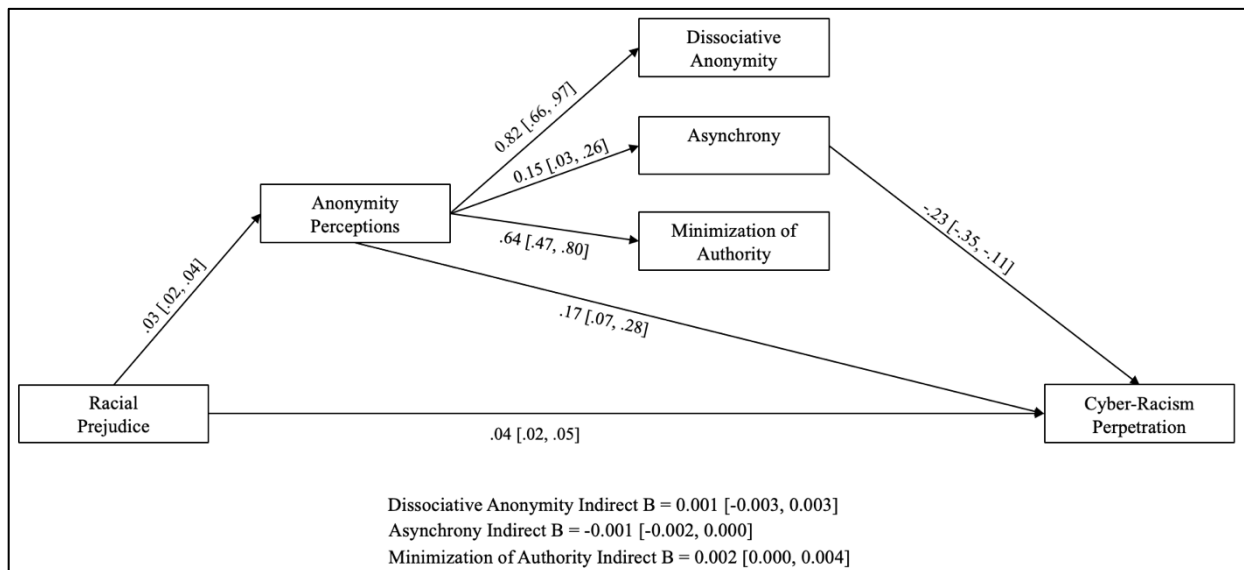


Figure 24. High Sadism Path Model

Chapter 4 - General Discussion

The current research investigated how the Dark Tetrad influences the integrated prejudice and cyber-racism theory (Studies 1 and 2) and the different dimensions of online disinhibition (Study 2). Victims of cyber-racism face many negative effects; anxiety, depression, and stress (Keum & Miller, 2017), and, therefore, understanding the processes specific to cyber-racism perpetration is needed, and an integrated theoretical model is desperately needed. By using theoretical explanations from the prejudice literature, the cyber-racism literature, and the online communication literature, the current research provided support for these processes and theoretical underpinnings. Given that the JSM (Crandall & Eshleman, 2003) and the Keum and Miller (2018) theorized model have been supported (White & Crandall, 2017; Barlett & Scott, 2023), this program of research examined how additional variables (i.e., Dark Tetrad, and dimensions of online disinhibition) impact the integrated theory. The current work was the first to (i) combine theories from prejudice literature and cyber-racism literature into an integrated

model to gain a more holistic understanding of cyber-racism perpetration (ii) test how individual differences moderate the mediating pathways in the integrated cyber-racism model and (iii) investigate how different online disinhibition dimensions impact cyber-racism perpetration.

Integrated Cyber-Racism Theory

To perform a racist action, one must hold some racial prejudice towards an outgroup. Thus, to understand the specific processes of cyber-racism perpetration, theories must include the importance of racial prejudice's role in said behaviors. The Justification-Suppression Model of Prejudice (JSM; Crandall & Eshleman, 2003) posits that people attempt to suppress their prejudice thoughts in order to appear socially acceptable. Suppression uses a large amount of energy, and the prejudice thoughts need to be released in a way that does not damage one's reputation. By providing adequate justification for one's thoughts, an individual may then express the normally socially unacceptable thoughts. Research has shown the JSM to explain expressions of prejudice towards a variety of groups, including racial outgroups (White & Crandall, 2017).

Current cyber-racism theory relies on specific online situations to explain why perpetration of racism happens in these spaces. While online, people may perceive themselves to be anonymous and, therefore, act differently than they would in a normal setting – online disinhibition (Keum & Miller, 2018). These anonymity perceptions and online disinhibition could act as a release of the usual suppression of prejudice thoughts – consistent with JSM theorizing, which could further explain cyber-racism perpetration. One of the goals of the current research was to test this integrated cyber-racism model that includes racial prejudice as the initial step in perpetrating cyber-racism.

Both Studies 1 and 2 provided additional support and replication of the original Keum and Miller (2017) model of cyber-racism such that high scores of anonymity were associated with high scores of online disinhibition which in turn was related to high scores of cyber-racism. Furthermore, Studies 1 and 2 supported the hypotheses that anonymity and online disinhibition partially mediated the relationship between racial prejudice and cyber-racism perpetration. In other words, higher scores of racial prejudice were associated with higher scores of anonymity and online disinhibition which in turn were related to high scores of cyber-racism. It should be noted that high scores of racial prejudice were directly associated with higher scores of cyber-racism as well. The results support the inclusion of prejudice within an integrated cyber-racism theory and hold important theoretical implications. It is difficult to imagine a racist action being taken without underlying racial prejudice, and these results suggest that there are benefits to including prejudice in theories and models specific towards an expression of prejudice behavior. In these studies specifically, racial prejudice was partially related to cyber-racism which suggests that it should not be excluded from discussions regarding the reasons people are racist online. Some research suggests treating online behavior as separate from in-person behavior, but by creating this separation, important non-online specific influences may be excluded which could provide a broader understanding of online behavior.

Individual Differences

Research has identified the Dark Tetrad as a predictor in a variety of online aggression like, cyberbullying, cyber-trolling, and cyber stalking (Brown et al., 2019; Furian & March, 2023; March et al., 2020; Xu et al., 2024). There have been conflicting results on which of the Dark Tetrad variables predict online aggression and additional research was needed to identify these relationships. A second goal of the current research was to expand our current

understanding of who the integrated cyber-racism model works for and if different Dark Tetrad traits influence a variety of relationships. The hypotheses that psychopathy, sadism, and the overall Dark Tetrad scores would moderate the various mediation relationships of the integrated cyber-racism theory were partially supported. Specifically, the integrated cyber-racism model worked for those who had high psychopathy and sadism scores compared to those who had low scores. Though no direct hypotheses were made regarding Machiavellianism and narcissism's ability to moderate these mediations, exploratory results suggest that the integrated cyber-racism model worked for those who had high Machiavellianism and narcissism scores compared to those who had low scores.

Interestingly, there were differences in the direct relationships that the Dark Tetrad traits moderated. For example, in Study 1, sadism moderated multiple direct relationships, such as the relationships between racial prejudice and cyber-racism. However, psychopathy did not moderate this direct relationship and only moderated two direct relationships within the model. It is unclear why different traits moderated varying direct relationships within this model, and additional research should investigate these individual relationships. Initial extrapolation may suggest that since sadism moderated the largest number of direct relationships (see Figure X), it may be a dominant trait important to the expression this racial prejudice.

The primary theoretical implications of this research pertain to the understanding of who may be more susceptible to the release of suppression provided by online factors. To our knowledge, no published research has investigated which Dark Tetrad traits moderate cyber-racism specifically. Though there has been support in cyberbullying literature for sadism and psychopathy as unique predictors of online harm (Van Geel et al., 2017; Goodboy & Martin, 2015), support for narcissism (Alavi et al., 2022; Goodboy & Martin, 2015) and

Machiavellianism (Alavi et al., 2022; Giumetti et al., 2022) as predictors of online harm has been mixed at best. Results from Study 1 add to the understanding of which traits increase the likelihood of cyber-racism. Study 1 found that all four Dark Tetrad traits moderated the integrated cyber-racism model such that the model only worked for those high in each Dark Tetrad traits compared to those who scored low. This suggests that in the process of expressing one's racial prejudice online, people who score higher on "darker" traits may be more susceptible to online racism when the social threat is diminished. The practical importance of these relationships underscores the need for replicability, and, therefore, Study 2 tested similar relationships albeit with a different measure of online disinhibition. Results from Study 2 did not find that any evidence that the Dark Tetrad traits moderated the integrated cyber-racism model. Though an interesting deviation, this can potentially be explained by the breakdown of online disinhibition and how the different dimensions impacted the integrated cyber-racism model.

Dimensions of Online Disinhibition

Suler (2004; 2005) identified online disinhibition as the act of behaving differently online than one would in a real-world situation. In the Keum and Miller (2018) model, online disinhibition acts as a mediator between anonymity and cyber-racism which was further integrated as a serial mediator in the current research. Both Study 1 and Study 2 supported the Keum and Miller (2018) model as online disinhibition mediated the relationship between anonymity and cyber-racism. Additionally, both studies supported general online disinhibition within the integrated cyber-racism model as both anonymity and online disinhibition mediated the relationship between racial prejudice and cyber-racism. To further the understanding of how online disinhibition mediates these relationships, Study 2 investigated if there were specific dimensions of online disinhibition that mediated these relationships more than others.

Little research has investigated how the different dimensions of online disinhibition (Suler, 2004; 2005) impact cyber-aggression. The Study 2 hypotheses focused on three dimensions that have been shown to predict other forms of online behavior; dissociative anonymity (Mueller-Coyne et al., 2022), asynchrony (Wang et al., 2024), and minimization of authority (Bertram, 2025), however, the other three dimensions were included in the analyses in an exploratory manner. The results from Study 2 did not support our hypotheses regarding the separate components of online disinhibition. Most moderated mediation results were not significant, except for the small negative relationship found with asynchrony. These results largely diverge from the results of Study 1 and the replicated results of Study 2 when the general online disinhibition was included in the model instead of the individual components. These findings hold important theoretical implications, as this suggests that measuring the components of online disinhibition separately is inappropriate within this model. Importantly, a composite online disinhibition assessment, rather than components, provides a more parsimonious explanation of the underlying theoretical tenets of the integrated model. Therefore, we recommend and posit that an integrated model that conceptualizes online disinhibition as more holistic better fits the data.

Limitations & Future Directions

Like all psychological research, limitations exist that require future research. First, although the current research focused on the Dark Tetrad and how these traits are related to the integrated cyber-racism theory, this is not an exhaustive list of individual differences that could impact cyber-racism. Research has identified a host of personality traits, emotional skills, and situational predictors of online aggression such as, the Big 5 (van Geel et al., 2017), coping skills (Kokkinos et al., 2013), and sleep quality (Kircaburun & Tosuntaş, 2018). Future research should

continue to test various predictors within the integrated cyber-racism model to provide a more encompassing understanding of why people are racist online.

Though the analyses in the current research provide a parsimonious interpretation of complex processes, it should be noted that the way in which the Dark Tetrad were grouped using a median split is a limitation. A median split holds multiple problems in that there may not be large differences in the cases around the median which could impact the groups they are assigned. In attempting to avoid using a median split, groups were initially split using standard deviations; however, the lack of variance in the predictors for those in the low (-1SD) classification led to non-convergence within the model. The lack of variance for participants in this group makes some sense; those low on sadism, for example, are likely not to perpetrate cyber-racism. Future research should use the Dark Tetrad as continuous moderators on simpler models to provide more parsimonious and meaningful results.

One of the main limitations of the current research is that it is correlational in nature, and thus, causality cannot be claimed through these studies. Correlational research is an important first step in furthering theory and thus these studies are needed to identify the general relationships between the Dark Tetrad and cyber-racism theory. The placement of the variables in the current research are based on the temporal ordering provided by theory (Keum & Miller, 2018) and the argument that to express prejudice, one must hold a form of those prejudice beliefs is consistent with JSM (Crandall & Eshleman, 2003). The Dark Tetrad was found to have an influence on the processes germane to cyber-racism, which suggests that future research should test these relationships longitudinally. Longitudinal support of these relationships would be a necessary step to determining causality. Moreover, experiments allow for causal claims; however, there is currently no valid way to measure cyber-racism in the moment, which should

be the goal of future research as this would allow research to focus on the causal relationships regarding cyber-racism.

Implications

Despite these limitations, the current research has important theoretical and practical implications. In terms of theoretical implications, these two studies investigated how individual differences in personality (the Dark Tetrad) influence an integrated cyber-racism model that pulled from prejudice research (Crandall & Eshleman, 2003) and specific cyber-racism theory (Kuem & Miller, 2018). Both Study 1 and Study 2 suggest that incorporating prejudice in models assessing cyber-racism is an important component that should not be excluded in the online world. Additionally, Study 2 contributed important information into how the dimensions of online disinhibition work in the case of cyber-racism theory. The results of Study 2 suggest that online disinhibition may be best measured as a whole construct and that deconstructing the dimensions does not add to our knowledge of how online disinhibition plays a role in cyber-racism processes.

Additionally, the current work has important practical implications in terms of identifying ways to make social media sites safer. As these results suggest that anonymity and online disinhibition play a role in the release of suppression for individuals high on the Dark Tetrad traits, online safety courses should aim to reduce anonymity perceptions which should in turn reduce one's online disinhibition. Through educating individuals on the processes in place to identify people online, we may be able to reduce this behavior from occurring. While no intervention study has examined the role of anonymity or online disinhibition in cyber-racism prevention and reduction, the You're Not Anonymous intervention includes curriculum aimed at teaching others that anonymity online is a myth, and results showed validity for cyberbullying

reduction (Barlett et al., 2020). However, again, no study has examined cyber-racism reduction. Results from Studies 1 and 2 suggest that this, and perhaps other, interventions may be successful at reducing cyber-racism, and future research should examine such effects. Moreover, it is unlikely that social media sites would put into place safeguards to limit the amount of anonymity one may use online, but these should still be taken into consideration in how accounts are allowed to be set up.

Conclusion

The current studies examined how individual differences and dimensions of online disinhibition impact an integrated cyber-racism theory. Specifically, this research investigated how the Dark Tetrad moderates an integrated prejudice and cyber-racism model (Studies 1 and 2) and if the dimensions of online disinhibition differ in their ability to mediate the relationships in the cyber-racism theory (Study 2). These studies provided both important theoretical implications for understanding and extending the current cyber-racism theory as well as practical implications in terms of who is most likely to perpetrate cyber-racism and in what situations. Overall, these studies provided a better understanding of the situations and individual differences that allow prejudice to be expressed without fear of social consequences.

References

- Alavi, M., Latif, A. A., Ramayah, T., & Tan, J. Y. (2022). Dark tetrad of personality, cyberbullying, and cybertrolling among young adults. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, <https://doi.org/10.1007/s12144-022-03892-4>.
- Arcila Calderón, C., Sánchez Holgado, P., Gómez, J., Barbosa, M., Qi, H., Matilla, A., Amado, P., Guzmán, A., López-Matías, D., & Fernández-Villazala, T. (2024). From online hate speech to offline hate crime: the role of inflammatory language in forecasting violence against migrant and LGBT communities. *Humanities & Social Sciences Communications*, 11(1), 1369. <https://doi.org/10.1057/s41599-024-03899-1>.
- Arnoult, M., Selimbegović, L., Hirschberger, G., Pyszczyński, T., & Chatard, A. (2023). Under the veil of tolerance: A justification–suppression approach to anti-Islamic implicit bias in reaction to terrorist attacks. *Journal of Personality and Social Psychology*, 125(2), 237. <https://doi.org/10.1037/pspa0000337>.
- Back, L. (2002). Aryans reading adorno: Cyber-culture and twenty-first century racism. *Ethnic and Racial Studies*, 25(4), 628–651. <https://doi.org/10.1080/01419870220136664>.
- Balta, S., Jonason, P., Denes, A., Emirtekin, E., Tosuntaş, Ş. B., Kircaburun, K., & Griffiths, M. D. (2019). Dark personality traits and problematic smartphone use: The mediating role of fearful attachment. *Personality and Individual Differences*, 149, 214-219. <https://doi.org/10.1016/j.paid.2019.06.005>.
- Barlett, C. P. (2015). Anonymously hurting others online: The effect of anonymity on cyberbullying frequency. *Psychology of Popular Media Culture*, 4(2), 70-79. <https://doi.org/10.1037/a0034335>.

- Barlett, C. P., & Gentile, D. A. (2012). Attacking others online: The formation of cyberbullying in late adolescence. *Psychology of Popular Media Culture*, 1(2), 123-135.
<https://doi.org/10.1037/a0028113>.
- Barlett, C. P., Heath, J. B., Madison, C. S., DeWitt, C. C., & Kirkpatrick, S. M. (2020). You're not anonymous online: The development and validation of a new cyberbullying intervention curriculum. *Psychology of popular media*, 9(2), 135. <https://doi/10.1037/ppm0000226>.
- Barlett, C. P., & Scott, J. E. (2023). Racism behind the screen: Examining the mediating and moderating relationships between anonymity, online disinhibition, and cyber-racism. *Journal of Personality and Social Psychology*, 125(6), 1332-1350. <https://doi/10.1037/pspa0000360>.
- Bertram, A. C. (2025). *How Rude: Expression of Incivility and Online Disinhibition in the Remote Workplace* (Unpublished Doctoral Dissertation, Capella University).
<https://www.proquest.com/openview/ccdc90b11c92afa5cc16a17fa31606f2/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y>.
- Branson, M., & March, E. (2021). Dangerous dating in the digital age: Jealousy, hostility, narcissism, and psychopathy as predictors of Cyber Dating Abuse. *Computers in Human Behavior*, 119, 8. <https://doi.org/10.1016/j.chb.2021.106711>.
- Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology*, 23, 1933-1967. <https://doi.org/10.1111/j.1559-1816.1993.tb01074.x>
- Brown, W. M., Hazraty, S., & Palasinski, M. (2019). Examining the Dark Tetrad and its links to cyberbullying. *Cyberpsychology, Behavior, and Social Networking*, 22(8), 552-557.
<https://doi.org/10.1089/cyber.2019.0172>.

- Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013) Behavioral Confirmation of Everyday Sadism. *Psychological Science*, 24(11), 2201–2209.
<https://doi.org/10.1177/0956797613490749>.
- Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment*, 83(1), 29–45. https://doi.org/10.1207/s15327752jpa8301_04.
- Campbell, W. K., Bush, C. P., Brunell, A. B., & Shelton, J. (2016). Understanding the social costs of narcissism: The case of the tragedy of the commons. *Personality and Social Psychology Bulletin*, 31(10), 1358–1368. <https://doi.org/10.1177/0146167205274855>.
- Capozza, D., Colledani, D., & Falvo, R. (2019). The Dark Triad, humanity attributions, and behavioral inclinations toward outgroups. *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, 26(3), 463-479. <https://doi.org/10.4473/TPM26.3.10>.
- Chabrol, H., Van Leeuwen, N., Rodgers, R., & Séjourné, N. (2009) Contributions of psychopathic, narcissistic Machiavellian, and sadistic personality traits to juvenile delinquency. *Personality and Individual Differences*, 47(7), 734-739.
<https://doi.org/10.1016/j.paid.2009.06.020>.
- 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 and Data Systems*, 121(1), 48-64. <http://dx.doi.org/10.1108/IMDS-08-2020-0509>.
- Christie, R., & Geis, F. L. (1970). Studies in Machiavellianism. *New York: Academic Press*.

- Chu, X., Li, Q., Fan, C., & Jia, Y. (2023). Life stress and cyberbullying: Examining the mediating roles of expressive suppression and online disinhibition. *Journal of Youth and Adolescence*, 52(8), 1647-1661. <https://doi.org/10.1007/s10964-023-01791-w>.
- Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, 129(3), 414. <https://doi.org/10.1037/0033-2909.129.3.414>.
- Crandall, C.S. & Eshleman, A. (2005). The Justification-Suppression Model of Prejudice : An Approach to The History of Prejudice Research. In Crandall, C.S. and Schaller, M. (Eds.) *Social Psychology of Prejudice: Historical and Contemporary Issues*. Lawrence, KS: Lewinian Press.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18. <https://doi.org/10.1037/0022-3514.56.1.5>
- Dixon, S. J. (2025, June 19). *Average daily time spent on social media worldwide 2012-2025*. Statista. <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/?srsltid=AfmBOoq-QhakOg9tR37PjD6AmUR-2e3wmJqHiFExVAiwCIkyjI4DNHBb>
- Duggan, M. (2017). *1 in 4 black Americans have faced online harassment because of their race or ethnicity*. Retrieved from www.pewresearch.org/fact-tank/2017/07/25/1-in-4-black-americans-have-faced-online-harassment-because-of-their-race-or-ethnicity [online resource] on 10/21/2024.

- Duncan, Z. & March, E. (2019). Using Tinder to start a fire: Predicting antisocial use of Tinder with gender and the Dark Tetrad. *Personality and Individual Differences*, 145, 9-14.
<https://doi.org/10.1016/j.paid.2019.03.014>.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- Furian, L., & March, E. (2023). Trolling, the Dark Tetrad, and the four-facet spectrum of narcissism. *Personality and Individual Differences*, 208, 1-6.
<https://doi.org/10.1016/j.paid.2023.112169>.
- Gajda, A., Moroń, M., Królik, M., Małuch, M., & Mraczek, M. (2022). The Dark Tetrad, cybervictimization, and cyberbullying: The role of moral disengagement. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*. <https://doi.org/10.1007/s12144-022-03456-6>.
- Gibb, Z. G., & Devereux, P. G. (2014). Who does that anyway? Predictors and personality correlates of cyberbullying in college. *Computers in Human Behavior*, 38, 8-16.
<https://doi.org/10.1016/j.chb.2014.05.009>.
- 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>.
- 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>.

- Greitemeyer, T. (2022). Honesty-humility, the dark tetrad, and ideological beliefs: their incremental validity in predicting explicit prejudice toward asylum seekers. *Personality and Individual Differences*, 197, 111786. <https://doi.org/10.1016/j.paid.2022.111786>.
- Grimm, P. (2010). Social Desirability Bias. In Wiley International Encyclopedia of Marketing (eds J. Sheth and N. Malhotra). <https://doi.org/10.1002/9781444316568.wiem02057>.
- Hare, R. D. (1985). Comparison of procedures for the assessment of psychopathy. *Journal of Consulting and Clinical Psychology*, 53, 7–16. <https://doi/10.1037/0022-006X.53.1.7>.
- Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Guilford publications.
- Hernández, M. P., Schoeps, K., Maganto, C., & Montoya-Castilla, I. (2021). The risk of sexual-erotic online behavior in adolescents—Which personality factors predict sexting and grooming victimization? *Computers in Human Behavior*, 114, 10. <https://doi.org/10.1016/j.chb.2020.106569>.
- Hoareau, N., Bagès, C., Allaire, M., & Guerrien, A. (2019). The role of psychopathic traits and moral disengagement in cyberbullying among adolescents. *Criminal Behaviour and Mental Health*, 29(5-6), 321-331. <https://doi.org/10.1002/cbm.2135>.
- Hoffarth, M. R., Hodson, G., & Molnar, D. S. (2018). When and why is religious attendance associated with antigay bias and gay rights opposition? A justification-suppression model approach. *Journal of Personality and Social Psychology*, 115(3), 526. <https://doi.org/10.1037/pspp0000146>.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

- Jonason, P. K., Underhill, D., & Navarrate, C. D. (2020). Understanding prejudice in terms of approach tendencies: The Dark Triad traits, sex differences, and political personality traits. *Personality and Individual Differences, 153*.
<https://doi.org/10.1016/j.paid.2019.109617>.
- Jones, D. N. (2013). Psychopathy and Machiavellianism predict differences in racially motivated attitudes and their affiliations. *Journal of Applied Social Psychology, 43*, E367-E378.
<https://doi.org/10.1111/jasp.12035>.
- Kaakinen, M., Sirola, A., Savolainen, I., & Oksanen, A. (2020). Impulsivity, internalizing symptoms, and online group behavior as determinants of online hate. *PLoS One, 15*(4).
<https://doi.org/10.1371/journal.pone.0231052>.
- Kaylin, N. (2024, January 8). *Black Mirror or Black Hole? American Phone Time Statistics*. Healthcare Data Management Software & Services | Harmony Healthcare IT. Retrieved from <https://www.harmonyhit.com/phone-screen-time-statistics/?clreqid=a5b99b78-5919-4ddd-a2d7-5cb3c4e66e6d&kbid=58587> [online resource] on 11/12/24
- Keum, B. T. (2021). Development and validation of the Perceived Online Racism Scale short form (15 items) and very brief (six items). *Computers in Human Behavior, 3*, 100082.
<https://doi.org/10.1016/j.chbr.2021.100082>.
- Keum, B. T. & Ahn, L. H. (2021). Impact of online racism on psychological distress and alcohol use severity: Testing ethnic-racial socialization and silence about race as moderators. *Computers in Human Behavior, 120*, 106773. <https://doi.org/10.1016/j.chb.2021.106773>.
- Keum, B. T., & Li, X. (2022). Online Racism, Rumination, and Vigilance: Impact on Distress, Loneliness, and Alcohol Use. *The Counseling Psychologist, 51*(3), 422-448. <https://doi.org/10.1177/00110000221143521>.

- Keum, B. T. & Miller, M. J. (2017). Racism in digital era: Development and initial validation of the Perceived Online Racism Scale (PORS v1.0). *Journal of Counseling Psychology*, 64(3), 310-324. <https://doi.org/10.1037/cou0000205>.
- Keum, B. T. & Miller, M. J. (2018). Racism on the Internet: Conceptualization and recommendations for research. *Psychology of Violence*, 8(6), 782-791. <http://dx.doi.org/10.1037/vio0000201>.
- King, E. B., Shapiro, J. R., Hebl, M. R., Singletary, S. L., & Turner, S. (2006). The stigma of obesity in customer service: A mechanism for remediation and bottom-line consequences of interpersonal discrimination. *Journal of Applied Psychology*, 91(3), 579-593. <https://doi.org/10.1037/0021-9010.91.3.579>
- Kırcaburun, K., & Tosuntaş, Ş. B. (2018). Cyberbullying perpetration among undergraduates: Evidence of the roles of chronotype and sleep quality. *Biological Rhythm Research*, 49(2), 247-265. <https://doi.org/10.1080/02723646.2017.1352918>.
- Kocaturk, M., & Bozdag, F. (2020). Xenophobia among University Students: Its Relationship with Five Factor Model and Dark Triad Personality Traits. *International Journal of Educational Methodology*, 6(3), 545-554. <https://doi.org/10.12973/ijem.6.3.545>.
- Koehn, M. A., Jonason, P. K., & Davis, M. D. (2019). A person-centered view of prejudice: The Big Five, Dark Triad, and prejudice. *Personality and Individual Differences*, 139, 313-316. <https://doi.org/10.1016/j.paid.2018.11.038>.
- Kokkinos, C. M., Antoniadou, N., Dalara, E., Koufoglazou, A., & Papatziki, A. (2013). Cyberbullying, personality and coping among pre-adolescents. *International Journal of Cyber Behavior, Psychology and Learning*, 3(4), 55-59. <https://doi.org/10.4018/ijcbpl.2013100104>.

- 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>.
- Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2), 434-443. <https://doi.org/10.1016/j.chb.2011.10.014>.
- Lapidot-Lefler, N., & Barak, A. (2015). The benign online disinhibition effect: Could situational factors induce self-disclosure and prosocial behaviors? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 9(2). <https://doi.org/10.5817/CP2015-2-3>.
- Lieberman, A., & Schroeder, J. (2020). Two social lives: How differences between online and offline interaction influence social outcomes. *Current Opinion in Psychology*, 31, 16-21. <https://doi.org/10.1016/j.copsyc.2019.06.022>
- Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment*, 66, 488–524. https://doi.org/10.1207/s15327752jpa6603_3.
- Machackova, H., Blaya, C., Bedrosova, M., Smahel, D., & Staksrud, E. (2020). Children's experiences with cyberhate. *EU Kids Online*. <https://doi.org/10.21953/lse.zenkg9xw6pua>.
- March, E., Litten, V., Sullivan, D. H., & Ward, L. (2020). Somebody that I (used to) know: Gender and dimensions of dark personality traits as predictors of intimate partner cyberstalking. *Personality and Individual Differences*, 163(6). <https://doi.org/10.1016/j.paid.2020.110084>.
- March, E., McDonald, L., & Forsyth, L. (2024). Personality and internet trolling: A validation study of a representative sample. *Current Psychology: A Journal for Diverse Perspectives*

- on *Diverse Psychological Issues*, 43(6), 4815-4818. <https://doi.org/10.1007/s12144-023-04586-1>.
- March, E., & Steele, G. (2020). High esteem and hurting others online: Trait sadism moderates the relationship between self-esteem and Internet trolling. *Cyberpsychology, Behavior, and Social Networking*, 23(7), 441-446. <https://doi.org/10.1089/cyber.2019.0652>.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio and S. L. Gaertner (Eds.), *Prejudice, Discrimination, and Racism* (p. 91-124). Academic Press Publishers.
- Mededović, J., & Petrovic, B. (2015). The Dark Tetrad. *Journal of Individual Differences*, 36(4), 228-236. <https://doi.org/10.1027/1614-0001/a000179>.
- Mueller-Coyne, J., Voss, C., & Turner, K. (2022). The impact of loneliness on the six dimensions of online disinhibition. *Computers in Human Behavior Reports*, 5, 100169. <https://doi.org/10.1016/j.chbr.2022.100169>.
- Nesdale, D. (2007). The development of ethnic prejudice in early childhood: Theories and research. In O. N. Saracho & B. Spodek (Eds.), *Contemporary Perspectives on Socialization and Social Development in Early Childhood Education* (pp. 213–240). IAP Information Age Publishing.
- Nitschinsk, L., Tobin, S. J., & Vanman, E. J. (2022). The disinhibiting effects of anonymity increase online trolling. *Cyberpsychology, Behavior, and Social Networking*, <https://doi.org/10.1089/cyber.2022.0005>.
- Nocera, T. R., Dahlen, E. R., Mohn, R. S., Leuty, M. E., & Batastini, A. B. (2022). Dark personality traits and anger in cyber aggression perpetration: Is moral disengagement to blame? *Psychology of Popular Media*, 11(1), 24-34. <https://doi.org/10.1037/ppm0000295>.

- O'Meara, A., Davies, J., & Hammond, S. (2011). The psychometric properties and utility of the Short Sadistic Impulse Scale (SSIS). *Psychological Assessment*, 23(2), 523–531. <https://doi.org/10.1037/a0022400>.
- Pabian, S., De Backer, C. J. S., & Vandebosch, H. (2015). Dark Triad personality traits and adolescent cyber-aggression. *Personality and Individual Differences*, 75, 41-46. <https://doi.org/10.1016/j.paid.2014.11.015>.
- Park, M. S., Billieux, J., Raj, S., Lee, M. C., Geoffrey, D. S., & Nuyens, F. (2024). Functional and dysfunctional impulsivity mediates the relationships between ‘Dark Triad’ traits and cyberbullying perpetration. *Criminal Behaviour and Mental Health*, 34(1), 54-65. <https://doi.org/10.1002/cbm.2321>.
- Paulhus, D. L., Buckels, E. E., Trapnell, P. D., & Jones, D. N. (2021). Screening for dark personalities: The Short Dark Tetrad (SD4). *European Journal of Psychological Assessment*, 37(3), 208–222. <https://doi.org/10.1027/1015-5759/a000602>.
- Petrosyan, A. (2025, February 6). *Daily time spent online by users worldwide Q3 2024*. Statista. <https://www.statista.com/statistics/1380282/daily-time-spent-online-global/>
- Raskin, R., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological Reports*, 45, 590. <https://doi/10.2466/pr0.1979.45.2.590>.
- Reidy, D. E., Zeichner, A., & Seibert, L. A. (2011) Unprovoked Aggression: Effects of Psychopathic Traits and Sadism. *Journal of Personality*, 79(1), 75-100. <https://doi-org.er.lib.k-state.edu/10.1111/j.1467-6494.2010.00691>.
- Rieger, D., Kümpel, A. S., Wich, M., Kiening, T., & Groh, G. (2021). Assessing the Extent and Types of Hate Speech in Fringe Communities: A Case Study of Alt-Right Communities

- on 8chan, 4chan, and Reddit. *Social Media + Society*, 7(4).
<https://doi.org/10.1177/20563051211052906>.
- Rosseel Y (2012). “lavaan: An R Package for Structural Equation Modeling.” *Journal of Statistical Software*, 48(2), 1–36. <http://doi:10.18637/jss.v048.i02>.
- Saucier, D. (2002). Self-Reports of Racist Attitudes for Oneself and for Others. *Psychologica Belgica*, 42(1-2), 99-105. <https://doi.org/10.5334/pb.987>.
- Scott, R. A., Stuart, J., & Barber, B. L. (2022). What predicts online disinhibition? Examining perceptions of protection and control online and the moderating role of social anxiety. *Cyberpsychology, Behavior, and Social Networking*, 25(5), 294-300.
<https://doi.org/10.1089/cyber.2021.0250>.
- Seo, M., Lee, M., & Lee, J. (2023). Justification of Discrimination against People with Mental Illness in Republic of Korea. *Healthcare*, 11(8), 1195.
<https://doi.org/10.3390/healthcare11081195>.
- 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>.
- 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, 106534. <https://doi.org/10.1016/j.chb.2020.106534>.
- Stober, J. (2001). The Social Desirability Scale – 17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17(3), 222-232. <https://doi/10.1027/1015-5759.17.3.222>.

- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology and Behavior*, 7, 321–326.
<http://dx.doi.org/10.1089/1094931041291295>
- Suler, J. (2005). The online disinhibition effect. *International Journal of Applied Psychoanalytic Studies*, 2(2), 184-188. <https://doi.org/10.1002/aps.42>.
- van Geel, M., Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, 106, 231-235.
<https://doi.org/10.1016/j.paid.2016.10.063>.
- Vogels, E. A. (2022). *Teens and cyberbullying: 2022*. Retrieved from [pewresearch.org/internet/2022/12/15/teens-and-cyberbullying-2022/](https://www.pewresearch.org/internet/2022/12/15/teens-and-cyberbullying-2022/) [online resource] on 10/21/2024.
- Wachs, S., Bilz, L., Wettstein, A., Wright, M. F., Kansok-Dusche, J., Krause, N., & Ballaschk, C. (2022). Associations between witnessing and perpetrating online hate speech among adolescents: Testing moderation effects of moral disengagement and empathy. *Psychology of Violence*, 12(6), 371-381. <https://doi.org/10.1037/vio0000422>.
- Wachs, S., Mazzone, A., Milosevic, T., Wright, M. F., O'Higgins Norman, J., & Blaya, C. (2021). Online correlates of cyberhate involvement among young people from ten European countries: An application of the routine activity and problem behavior theory. *Computers in Human Behavior*, 123(3). <https://doi.org/10.1016/j.chb.2021.106872>.
- Wachs, S., & Wright, M. F. (2018). Associations between Bystanders and Perpetrators of Online Hate: The Moderating Role of Toxic Online Disinhibition. *International Journal of Environmental Research and Public Health*, 15(9).
<https://doi.org/10.3390/ijerph15092030>

- Wang, Y., Hawk, S. T., Wong, N., & Zhang, Y. (2023). Lonely, impulsive, and seeking attention: Predictors of narcissistic adolescents' antisocial and prosocial behaviors on social media. *International Journal of Behavioral Development*, 47(6), 540-547.
<https://doi.org/10.1177/01650254231198034>.
- Wang, L., Jiang, S., Zhou, Z., Fei, W., & Wang, W. (2024). Online disinhibition and adolescent cyberbullying: A systematic review. *Children and Youth Services Review*, 156, 107352.
<https://doi.org/10.1016/j.childyouth.2023.107352>
- Wang, L., & Ngai, S. S. (2020). The effects of anonymity, invisibility, asynchrony, and moral disengagement on cyberbullying perpetration among school-aged children in China. *Children and Youth Services Review*, 119, 9.
<https://doi.org/10.1016/j.childyouth.2020.105613>.
- White, M. H. II, & Crandall, C. S. (2017). Freedom of racist speech: Ego and expressive threats. *Journal of Personality and Social Psychology*, 113(3), 413-429.
<https://doi.org/10.1037/pspi0000095>
- Wirawan, G. B. S., Hanipraja, M. A., Chrysanta, G., Imtaza, N., Ahmad, K. T., Marlina, I., Mahendra, D., & Larosa, A. T. (2022). Anxiety and prior victimization predict online gender-based violence perpetration among Indonesian young adults during COVID-19 pandemic: cross-sectional study. *Egyptian Journal of Forensic Sciences*, 12(1).
<https://doi.org/10.1186/s41935-022-00292-4>.
- Wright, M. F. (2013). The relationship between young adults' beliefs about anonymity and subsequent cyber aggression. *Cyberpsychology, Behavior, and Social Networking*, 16(12), 858-862. <http://dx.doi.org/10.1089/cyber.2013.0009>.

- Wright, M. F. (2014). Predictors of anonymous cyber aggression: The role of adolescents' beliefs about anonymity, aggression, and the permanency of digital content. *Cyberpsychology, Behavior, and Social Networking*, 17(7), 431-438.
<https://doi.org/10.1089/cyber.2013.0457>.
- Wright, M. F., Wachs, S., Huang, Z., Kamble, S. V., Soudi, S., Bayraktar, F., Li, Z., Lei, L., & Chang, S. (2022). Longitudinal Associations among Machiavellianism, Popularity Goals, and Adolescents' Cyberbullying Involvement: The Role of Gender. *Journal of Genetic Psychology*, 183(5), 482-493. <https://doi.org/10.1080/00221325.2022.2095251>.
- Wu, S., Lin, T. C., & Shih, J. F. (2017). Examining the antecedents of online disinhibition. *Information Technology & People*, 30(1), 189-209. <https://doi/10.1108/ITP-07-2015-0167>.
- Xu, W., Zhao, B., & Jin, C. (2024). A meta-analysis of the relationship between personality traits and cyberbullying. *Aggression and Violent Behavior*, 79.
<https://doi.org/10.1016/j.avb.2024.101992>
- Zhong, L. R., Keibell, M. R., & Webster, J. L. (2020). An exploratory study of Technology-Facilitated Sexual Violence in online romantic interactions: Can the Internet's toxic disinhibition exacerbate sexual aggression? *Computers in Human Behavior*, 108, 10.
<https://doi.org/10.1016/j.chb.2020.106314>.

Appendix A
Short Dark Triad Scale (SD4; Paulhus et al., 2021)

Identity 1: Crafty (Machiavellianism)

1. It's not wise to let people know your secrets.
2. Whatever it takes, you must get the important people on your side.
3. Avoid direct conflict with others because they may be useful in the future.
4. Keep a low profile if you want to get your way.
5. Manipulating the situation takes planning.
6. Flattery is a good way to get people on your side.
7. I love it when a tricky plan succeeds.

Identity 2: Special (Narcissism)

8. People see me as a natural leader.
9. I have a unique talent for persuading people.
10. Group activities tend to be dull without me.
11. I know that I am special because people keep telling me so.
12. I have some exceptional qualities.
13. I'm likely to become a future star in some area.
14. I like to show off every now and then.

Identity 3: Wild (Psychopathy)

15. People often say I'm out of control.
16. I tend to fight against authorities and their rules.
17. I've been in more fights than most people of my age and gender.
18. I tend to dive in, then ask questions later.
19. I've been in trouble with the law.
20. I sometimes get into dangerous situations.
21. People who mess with me always regret it.

Identity 4: Mean (Sadism)

22. Watching a fistfight excites me.
23. I really enjoy violent films and video games.
24. It's funny when idiots fall flat on their face.
25. I enjoy watching violent sports.
26. Some people deserve to suffer.
27. Just for kicks, I've said mean things on social media.
28. I know how to hurt someone with words alone.

Appendix B

Anonymity Perceptions Scale (Wright, 2013)

1. By being anonymous, I would not fear that my mean online behaviors could lead me to being punished by authority figures, if I engaged in such behaviors.
2. By being anonymous, I do not fear my mean online behaviors could lead to retaliation by the target of the behaviors, if I engaged in such behaviors.
3. I am confident that I would not be caught if I engaged in mean online behaviors.

Appendix C

Measure of Online Disinhibition (MOD; Stuart & Scott, 2021)

1. I act differently online than I do offline
2. I act tougher on the internet than I do face-to-face
3. I am less cautious about what I say online than about what I say in person
4. I am more assertive online than I am offline
5. I am more competitive online than I am offline
6. I am more confident online than I am offline
7. I am more outgoing online than I am offline
8. I am more able to discuss controversial issues online than I am in person
9. I find communicating with others easier on the internet than in person
10. I make friends more easily online than I do offline
11. I say things on the internet that I would not say in person
12. My behaviors online are less restricted than in person

Appendix D

Social Desirability Scale – 17 (SDS-17; Stober, 2001)

1. I sometimes litter.
2. I always admit my mistakes openly and face the potential negative consequences.
3. In traffic I am always polite and considerate of others.
4. I have tried illegal drugs (for example, marijuana, cocaine, etc.).
5. I always accept others' opinions, even when they don't agree with my own.
6. I take out my bad moods on others now and then.
7. There has been an occasion when I took advantage of someone else.
8. In conversations I always listen attentively and let others finish their sentences.
9. I never hesitate to help someone in case of emergency.
10. When I have made a promise, I keep it--no ifs, ands or buts.
11. I occasionally speak badly of others behind their back.
12. I would never live off other people.
13. I always stay friendly and courteous with other people, even when I am stressed out.
14. During arguments I always stay objective and matter-of-fact.
15. There has been at least one occasion when I failed to return an item that I borrowed.
16. I always eat a healthy diet.
17. Sometimes I only help because I expect something in return.

Appendix E
Attitudes Towards Black People Scale (Brigham, 1993)

1. If a Black person were put in charge of me, I would not mind taking advice and direction from him or her.
2. If I had a chance to introduce Black visitors to my friends and neighbors, I would be pleased to do so.
3. I would rather not have Black people live in the same apartment building I live in.
4. I would probably feel somewhat self-conscious dancing with a Black person in a public place.
5. I would not mind it at all if a Black family with about the same income and education as me moved in next door.
6. I think that Black people look more similar to each other than White people do.
7. Interracial marriage should be discouraged to avoid the “who-am-I? confusion which the children feel.
8. I get very upset when I hear a White person make a prejudicial remark about Black people.
9. I favor open housing laws that allow more racial integration of neighborhoods.
10. It would not bother me if my new roommate was Black.
11. It is likely that Black people will bring violence to neighborhoods when they move in.
12. I enjoy a funny racial joke, even if some people might find it offensive.
13. The federal government should take decisive steps to override the injustices Black people suffer at the hands of local authorities.
14. Black and White people are inherently equal.
15. Black people are demanding too much too fast in their push for equal rights.
16. White people should support Black people in their struggle against discrimination and segregation.
17. Generally, Black people are not as smart as White people.
18. I worry that in the next few years I may be denied my application for a job or a promotion because of preferential treatment given to minority group members.
19. Racial integration (of schools, businesses, residences, etc.) has benefitted both White people and Black people.
20. Some Black people are so touchy about race that it is difficult to get along with them.

Appendix F

Online Racism Perpetration Scale (Barlett & Scott, 2023)

1. Sent others racist insults regarding their online profile (e.g., profile pictures, user ID).
2. Harassed another (e.g., trolled) by starting a racist argument about them for no reason.
3. Sent a racist meme (e.g., racist catchphrases, captioned photos, #hashtags etc.).
4. Tagged others in (or shared) racist content (e.g., websites, photos, videos, posts) insulting another's race/ethnicity.
5. Sent online posts with racist comments.

Appendix G
Online Disinhibition Scale (Cheung et al., 2020)

Dissociative Anonymity

1. I feel I am anonymous in the online environment.
2. I believe that my personal identity remains unknown to others in the online environment.
3. I feel that I can hide my identity online.
4. My actions are not identifiable in the online environment.

Invisibility

1. I feel invisible in the online environment.
2. I feel that online environments avert others from seeing any of me.
3. I am invisible online.
4. Others do not see me in the online environment.
5. My actions are invisible in the online environment.

Asynchronicity

1. I do not need to reply to others immediately in the online environment.
2. I can delay my feedback to others in the online environment.
3. I can postpone replying to others in the online environment.

Solipsistic Introjection

1. I assign a character to that person I am communicating with online.
2. I interpret others' messages with my expectation during online communication.
3. I perceive how that person's intended to talk in the online environment.

Dissociative Imagination

1. I feel like people in the online space are just imaginary with no connection to reality.
2. The online environment is an imaginary world.
3. The online environment has no connection to reality.
4. The online environment is separated from the offline world.

Minimization of Authority

1. I am away from real life authorities in the online environment.
2. I feel less fear of offline authorities in the online environment.
3. I feel free from real life authorities in the online environment.
4. I feel that offline authorities are absent in the online environment.