How Doxing on Social Media Leads to Social Stigma and Perceived Dignity

Guohou Shan
Temple University
guohou.shan@temple.edu

Wenxi Pu
University of Manitoba
wenxi.pu@umanitoba.ca

Jason Bennet Thatcher
Temple University
jason.thatcher@temple.edu

Philip Roth
Clemson University
rothp@clemson.edu

Abstract

Doxing is the practice of publicly posting someone else's personal information online (e.g., on social media) without their consent. Doxing on social media can damage the social image of doxing victims (doxees). Three types of doxing (deanonymizing, targeting, and delegitimizing) are documented in the literature. To better understand and mitigate the harmful consequences, in this research, we use social identity threat theory to propose a research model. We aim to understand whether doxing leads to social stigma and loss of perceived dignity, whether types of doxing result in different outcomes, and to unpack the mechanisms of doxing impact. Through two online experiments, we establish that doxing leads to social stigma and loss of perceived dignity; only delegitimizing doxing matters and perceived trustworthiness can help explain the effects of doxing on perceived dignity, but not on social stigma. Theoretical and practical implications are discussed in the paper.

Keywords: Doxing; social stigma; perceived dignity; social identity threat theory; perceived trustworthiness.

1. Introduction

Doxing refers to "the intentional public release onto the Internet of personal information about an individual by a third party, often with the intent to humiliate, threaten, intimidate, or punish the identified individual" (Douglas, 2016, pg.1). Doxing can occur for various reasons, including online shaming, hacktivism, and blackmail. Based on the purpose of doxing and the information doxed, doxing can be classified into three doxing types: deanonymizing doxing (Type I), targeting doxing (Type II), and delegitimizing doxing (Type III) (Li & Whitworth, 2023), which reveal information such as (I) name, (II) home address, and (III) involvement in "immoral" activities, respectively.

Doxing is related to but different from privacy violation, cyberbullying, naming and shaming, and digital vigilantism. Privacy violation is the behavior of unauthorized or improper access, use, or disclosure of an individual's personal information or data (Hann et al., 2007). Cyberbullying on social media refers to the aggressive, harmful behavior conducted by a group or an individual on social media repeatedly over time against targets who cannot defend themselves easily (Chan et al., 2022). Naming and shaming is a practice where individuals or entities are publicly identified and criticized for their actions, often with the goal of holding them accountable or encouraging behavioral change. (McDonald, 2020). Digital vigilantism is "a process where citizens are collectively offended by other citizen activity, and respond through coordinated retaliation on digital media, including mobile devices and social media platforms." (Trottier, 2017, pg. 56).

Doxing highlights the careful *curation* of doxee's information. That is, this information can be either public, but hard to access, or private. This makes doxing different from privacy violation, which involves private information only. Naming and shaming as well as cyberbullying tend to involve subjective "finger-pointing" but typically do not involve careful curation of the victim's personal information (West et al., 2014). Doxing differs from digital vigilantism because the doxee might not be offending to the collective. In sum, doxing is different from these related phenomena and can potentially be used as a "tool" to facilitate these phenomena.

Doxing can occur on many forums, but it is particularly pernicious and widespread on social media platforms such as Facebook, Twitter, and Instagram (Yudiana et al., 2022). Doxing has many negative effects. First, doxers can strip doxees' anonymity in order to cause harm (Douglas, 2016). Second, doxing home addresses makes the doxees vulnerable to physical harassment (Douglas, 2016). Third, doxing engagement in "immoral" activities can undermine the doxees' credibility and reputation (Douglas, 2016). Fourth, doxing the political affiliations can affect job hiring evaluation (Roth et al., 2023a; Roth et al., 2023b).

Doxing merits more attention in sociotechnical research, especially in information systems (IS) privacy and security research, because of its implications for the well-being of doxees. A recent literature review on doxing by Fang et al. (2023) calls

URI: https://hdl.handle.net/10125/106389
978-0-9981331-7-1

for more empirical investigations that lead to a richer theoretical understanding of doxing from behavioral perspectives. In line with this call, we investigate whether and how doxing leads to stigmatization of victims and poses threats to their dignity. Stigma is an attribute or mark that deeply discredits people in society (Goffman, 1963; Link & Phelan, 2001) and dignity is "the recognition that human beings possess intrinsic value and, as such, are endowed with certain rights and should be treated with respect" (Leidner & Tona, 2021, pg. 343)

Whether doxing impacts social stigma and perceived dignity is not intuitive. On the one hand, doxing engagement in "immoral" activities (i.e., delegitimizing doxing) can undermine the doxees' credibility and reputation (Douglas, 2016), threatening their dignity and leading to stigmatization. On the other hand, the doxees can receive compassion and sympathy due to their being doxed (Tan, 2022), which may reduce doxees' likelihood of being stigmatized and threatened on their dignity.

Furthermore, understanding the different impacts of doxing on doxees' social stigma and perceived dignity and the underlying mechanisms is important for social media platforms to design different mitigation strategies (e.g., proactive and reactive) to resolve the impact. Also, figuring out the differential impacts of three types of doxing can enrich the literature related to doxing. Therefore, we propose the following research questions.

RQ1: Does doxing lead to social stigma and perceived dignity?

RQ2: Do different types of doxing content matter in terms of impacts?

RQ3: What is the underlying mechanism of the doxing impact?

By answering these questions on doxing, stigma, and dignity, this study sets out to achieve three main objectives. First, we aim to establish the causal impact of doxing on social stigma and perceived dignity. Second, we seek to clarify our understanding of whether and how different types of doxing content may have different effects. Third, we strive to shed light on the underlying mechanism of the impact of doxing on social stigma and perceived dignity. We draw upon social identity threat theory to develop our research model. We empirically test the proposed research model using online experiments.

This study provides several theoretical contributions. First, it enriches the IS literature by empirically investigating doxing on social media and its impact. Fang et al. (2023) shows a need for research on the impacts of doxing, such as how doxing may affect social stigma and dignity and whether different types of doxing content matter. Our research contributes to the understanding of the underlying mechanism of the impact of doxing on social stigma based on the social identity threat theory. Second, we also enrich the IS literature by investigating the phenomena of social stigma and dignity. IS researchers have started theoretically investigating how personal data digitization (PDD) interacts with human dignity (Leidner & Tona, 2021). Doxing has not been empirically studied as an antecedent to social stigma and perceived dignity. Third, we extend the generalizability of social identity threat theory to the context of social media doxing.

2. Literature review

2.1. Doxing

According to Fang et al. (2023), who reviewed 28 research studies from various fields, doxing has received little attention in the literature, especially empirically investigating the impact of doxing from behavioral perspectives. Currently, researchers have mainly used conceptualization, design science, and qualitative methods to study the phenomenon of doxing (e.g., Douglas, 2016), doxing intention (e.g., Snyder et al., 2017), outcomes of being doxed (e.g., Franz & Thatcher, 2023), doxing detection (e.g., Farnum & Karimi, 2020), and doxing mitigation strategy (e.g., Solo, 2019).

Our research focuses on enriching this stream of literature by investigating whether and how doxing on social media affects the doxees' social stigma and perceived dignity. Our research findings can help better understand the phenomenon and its outcomes and help design potential mitigation strategies.

2.2. Social stigma and dignity on social media platforms

A stigma can be viewed as a socially constructed label that is linked to a negatively valued stereotype and that results in the marginalization of the bearers of the label (Link & Phelan, 2001). Those who devalue or marginalize the bearers of the label are the stigmatizers, while the ones who bear such a label are stigmatized (Yang et al., 2007). Hence, social stigma is the social devaluation of the stigma bearers that are linked to a negatively valued stereotype.

A number of research studies have examined stigma on social media platforms. However, current studies mainly investigate stigmas such as physical and mental health issues, Covid-19, LGB (short for Lesbian, Gay, and Bisexual), pregnancy loss, infertile women, and poverty (e.g., Chopra & Arora, 2020;

Meese et al., 2020; Öztürk et al., 2021). For example, Chopra and Arora (2020) found that contracting Covid-19 led to social stigma. Meese et al. (2020) found that social media can be a channel to destigmatize poverty. Öztürk et al. (2021) found that infertile women experienced social stigma on social media platforms and were less likely to report it. However, whether and how doxing leads to social stigma is not well understood.

The dignity literature has mostly focused on proposing the conceptualization and theories around human dignity (Leidner & Tona, 2021; Leidner et al., 2021; Mason et al., 2021). For example, Leidner and Tona (2021) proposed CARE theory (short for Claims, Affronts, Response, Equilibrium) to connect personal data digitalization (PDD), which is defined as "the sociotechnical encounters associated with the digitization of personal data for use in digital technologies" (p. 343), and human dignity. Specifically, they explain how different types of PDDs relate to different types of human dignity (e.g., behavioral, meritocratic, and inherent dignity). Most relevant to our study, they argued that showing others (i.e., posting information about others, a form of PDD) can lead to dignity affront. In our case, doxing is to show others without consent, and it can potentially cause an affront to the doxees' dignity. We capture such affront in (lowered) perceived dignity. We evaluate such affronts by measuring others' perceptions about the doxee's dignity.

To sum up, there remains a need to empirically investigate doxing's impact on social stigma and perceived dignity. Our research provides empirical evidence of the impact of doxing on social stigma and perceived dignity.

3. Hypothesis Development

In this section, we develop our research hypotheses (shown in Figure 1), relying on social identity threat theory and CARE theory.

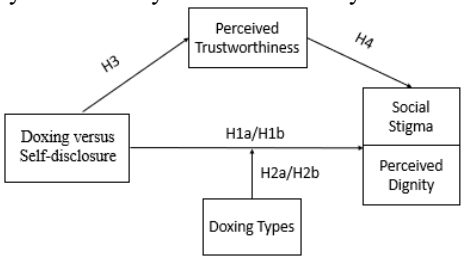

Figure 1. Research Model.

3.1. Information privacy and doxing

Doxing results from weaponization of information, which is a privacy issue (Lee, 2015) that has been widely investigated in the Information Systems (IS) field (Bélanger & Crossler, 2011; Mitra & Ransbotham, 2015). Information weaponization, especially deliberately using information with malintent, may lead to cyberattacks, cyberbullying, and online harassment (Lee, 2015; Mitra & Ransbotham, 2015) that induce social stigma and affront doxees' dignity. We choose these two constructs as the consequences of doxing based on social identity threat theory and CARE theory, which we will elaborate on in the next section.

3.2. Doxing's Impact on Stigma and Dignity

Social identity threat theory states that people belonging to a social group can be negatively affected and devalued if the group they belong to is threatened by the activation of negative group stereotypes (Martiny & Nikitin, 2019). Social identity threat theory depicts the concerns that people have about their positive image being threatened due to the activation of negative stereotypes of the groups they belong to (Martiny & Nikitin, 2019). Stigma can be used to evaluate the "negativeness" of specific stereotypes. Stigma can be viewed as a discrediting attribute (or mark) that is associated with intensive negative social evaluations, stigmatization can occur for one person when there is a deviation in their social image from socially accepted characteristics or stereotypes (Goffman, 1963).

Social identity threat theory, thus, leads us to use (perceived) social stigma as a potential consequence and helps to explain the link between doxing and social stigma. Steele et al. (2002) proposed that a person is stigmatized if their positive social image is threatened by the activation of negative group stereotypes. In the doxing context, the doxees' private information is involuntarily posted online, typically with malintent. This may mean that the doxees are powerless against the doxers. This can potentially produce the perception of powerlessness on the doxees. Also, doxing "immoral" activities can reveal the doxees' socially undesirable stereotypes (e.g., terrorists are violent and harmful to society) (Banko et al., 2020). Theoretically, according to social identity threat theory, the doxees can be stigmatized because of the activation of the negative stereotypes related to doxees and the potential negative doxing content.

CARE is a grand theory (Leidner and Tona 2021) that theorizes about the dynamics of human dignity in the context of personal data digitalization (PDD). Micro-level showing-others PDD encounters are of specific interest in this paper, and we will focus on discussing this type of PDD here (please see Leidner

& Tona, 2021 for a detailed introduction on other types of PDD). In this type of PDD encounters, doxers (the shower and posters) are trying to present a curated set of personal information about an individual (the doxee) to humiliate or harm the individual. CARE theory suggests dignity affront as a potential consequence of doxing. Due to the negative image brought by being doxed, the doxees may be less respected, which threatens their dignity. Such threats might penetrate different types of dignity. Therefore, we hypothesize:

Hypothesis H1a: Doxing on social media leads to higher social stigma.

Hypothesis H1b: Doxing on social media leads to lower perceived dignity.

Different types of doxing affect the social stigma and the perceived dignity differently. The first two types of doxing are related to deanonymizing and targeting, where the legal names and physical addresses are released, while the third type of doxing is related to delegitimizing, which involves releasing the immoral activities the doxees attend (Douglas, 2016). Compared with the third type, the first two types of doxing are less concerned with being associated with negative stereotypes, as suggested by Li and Whitworth (2022). Hence, their impact on the social stigma and the perceived dignity is less compared to the third type of doxing. Hence, we propose:

Hypothesis H2a: Different types of doxing content have different impacts on social stigma.

Hypothesis H2b: Different types of doxing content have different impacts on perceived dignity.

3.3. Perceived trustworthiness

We also aim to investigate the underlying mechanism to explain why doxing leads to social stigma and threats to perceived dignity. Trust is a fundamental element in our social and organizational context to facilitate smooth function of our societies and organizations (Mayer et al., 1995). Therefore, we focus on trustworthiness as a mediating mechanism of the doxing effect. Doxees typically bear the shadow of being viewed as powerless victims who could not protect their personal information. This can potentially taint evaluations of the doxees' trustworthiness. Doxing about engaging in "immoral" activities can produce a stereotype with negative social images, such as being violent or anti-social. Those negative stereotypes fundamentally decrease doxees' trustworthiness to be good citizens, as suggested by Curtis (2015). Hence, we select the perceived trustworthiness as a mechanism to explain the effect.

Stigma happens when a user is associated with negative social images and stereotypes (Goffman, 1963). When others perceive a person as less trustworthy, the person is potentially associated with negative images and stereotypes. This makes it more likely for them to be discriminated against and stigmatized. As trustworthiness can indicate credibility and strengthen dignity (Edlund et al., 2013), lower trustworthiness can lower perceived dignity. Hence, when the perceived trustworthiness is lower for a person, their dignity is more likely to be threatened. Therefore, we hypothesize:

Hypothesis 3 (H3): Doxing will reduce the doxee's perceived trustworthiness.

Hypothesis 4 (H4): The perceived trustworthiness will decrease the social stigma and increase perceived dignity.

4. Method

We conducted online experiments via Amazon Mechanical Turk (AMT) that manipulated sources of information (self or doxer) around the three types of doxing content (Types I - III). Because doxing is a type of disclosing information by others (Douglas, 2016), we used the self-disclosure of the same information as the control group for the doxing conditions. To create the manipulations, we first created a basic Facebook page. To increase the design's external validity and protect privacy, we borrowed elements from real users and used common American names for young adults. Several authors iterated on the manipulation processes several times to make the basic Facebook page look authentic and representative and to minimize distractions from our main manipulations.

For the manipulations, we followed Douglas (2016) to prepare real names, home and work addresses, and activities (we used capital riot activity because it was a real doxing example). For the doxing versus self-disclosure manipulations, we follow real-life examples of revealing doxing content via posts. To summarize, the experiment followed a two-by-three between-subject design (shown in Table 1) with factors of information source (doxing versus self-disclosure) and doxing types.

For the experimental flow, every participant first looked at the same Facebook user profile and then was randomly assigned to read an additional Facebook post coming from one manipulated experiment condition before evaluating the users' social stigma, perceived dignity, and perceived trustworthiness. After that, they completed manipulation checks.

Finally, the subject filled out demographic questions. To ensure the data quality, we included two attention checks ("Please tell us what animals are found on the moon: (a) dogs; (b) cats; (c) elephants;

(d) none" and "Please click 'Disagree', if you are paying attention") in our survey and two manipulation checks ("The additional Facebook post I examined was: (a) posted by the Facebook user themselves; (b) posted by another Facebook user; c) other" and "The content of the additional Facebook post is about (a) name, phone number, and e-mail address (b); home address and working address; (c) an activity attended d) other"). The basic Facebook page and the manipulated posts are shown in Figures 2 and 3.

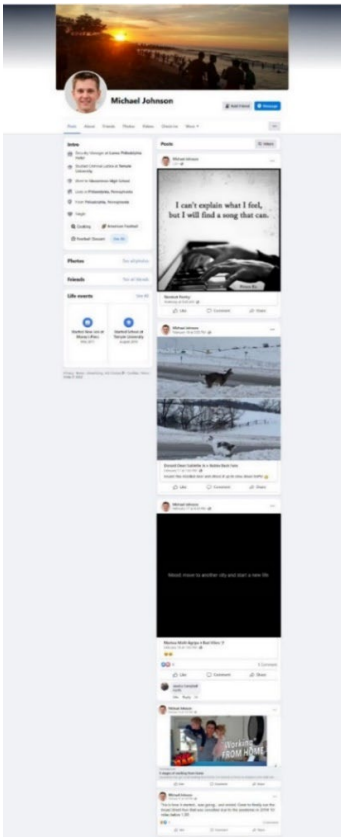

Figure 2. Designed Facebook profile.

Table 1. Two-by-three experimental design

Study Design		Doxing Types		
		Type I	Type II	Type III
Information Source	Doxing	Group 1	Group 2	Group 3
	Self-disclosure	Group 4	Group 5	Group 6

Figure 3(a). Manipulated posts (Group 1)

Figure 3(b). Manipulated posts (Group 2)

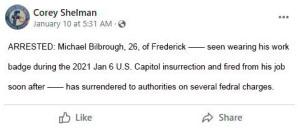

Figure 3(c). Manipulated posts (Group 3)

Figure 3(d). Manipulated posts (Group 4)

Figure 3(e). Manipulated posts (Group 5)

Figure 3(f). Manipulated posts (Group 6)

We follow previous literature to measure the social stigma, perceived dignity, and perceived trustworthiness. Specifically, for the social stigma, we adapted items from the Internalized Stigma of Mental Illness scale for evaluating the self-stigma from mental illness (Ritsher et al., 2003). We switched the self-oriented items to a third-person perspective to capture how others perceive the doxee. For perceived trustworthiness, we adapted items from Griskevicius et al. (2006). For perceived dignity, we measured three different dignity types by adapting the terms proposed by Leidner and Tona (2021).

5. Results

We first conducted one study to test the validity and reliability of our constructs and experimental flow by collecting 200 AMT responses. 181 participants completed our experiments. Then, based on the results, we improved the experimental design and conducted another study on another 100 AMT users. 96 participants completed that experiment.

5.1. Results of the first study

5.1.1. Data description and various checks. The sample characteristics for Study 1 appear in Table 2. Out of the 181 respondents, men accounted for 61.9%, and women made up 38.1%. Participants were predominantly college-educated or higher, were young to middle-aged (25-55), and supported the Democratic party (59.7%). Furthermore, for the two attention checks and two manipulation checks, the passing rate is 71.8% (130/181), 95.6% (173/181), 65.7% (119/181), and 61.9 % (112/181), respectively.

Table 2. Sample characteristics.

Variable	Frequency	Percentage (%)
Gender		
Men	112	61.9
Women	69	38.1
Education (degree)		
High school	7	3.9
Some college	8	4.4
Bachelor's degree	104	57.5
Master's degree	61	33.7
Doctorate	1	.5
Other	0	0
Age		
<25	20	11.0
26-35	99	54.7
36-45	30	16.6
46-55	24	13.3
56+	8	4.4
Political affiliation		
Democrats	108	59.7
Republicans	46	25.4
Other	27	14.9

5.1.2. Measures, reliability, and validity. Using the whole sample (n=181), we measure the construct reliability by both composite reliability (CR) and Cronbach alpha (Alpha). The results are shown in Table 3, which indicates high reliability. For construct validity, we conducted a confirmatory factor analysis (CFA) analysis, which suggested a good model fit. The overall model Satorra-Bentler chi-square statistic was 1734.484 with p =.000 (CFI = 0.893, RMSEA = 0.074, RMSEA intervals of 0.069, 0.079, and SRMR = 0.065) (Satorra & Bentler, 2001). Scale average variance extracted (AVEs) exceeded .70, indicating convergent validity. The square root of the AVE was higher than the latent construct correlations, suggesting discriminant validity (shown in Table 5). CFA, standardized loadings, and scale reliabilities appear in Table 3. We also put the summary statistics of variables used in this study in Table 4, where we can see that the mean values of social stigma, first, second, and third types of dignity, perceived trustworthiness, doxing, and doxing Type I, II, and III are 4.75, 5.16, 5.12, 5.67, 5.59, .47, .33, .33, and .35, respectively.

Table 3. Scale and item descriptions, loadings, standard error, and reliability.

Scale and item description	Factor loading	S.E.	CR	Alpha
Social stigma (SS) Tell us your opinion about this user in terms of the following:			.963	.963
This user tends to be violent.	.860***	.021		
This user shouldn't get married.	.770***	.032		
I feel this user is out of place in the world.	.763***	.032		
I discriminate against this user.	.876***	.019		
I feel ashamed for this user.	.869***	.020		
I feel superior to this user.	.792***	.029		
We should not socialize with this user.	.867***	.020		
We should avoid getting close to this user.	.900***	.016		
We should isolate this user.	.903***	.015		
We should stay away from this user.	.902***	.015		
Dignity 1 (DT1) Please let us know your opinion in terms of the following:			.943	.942
I think this user is benevolent	.844***	.026		
I think this user is righteous	.841***	.028		
I think this user has integrity	.843***	.024		
I think this user has wisdom	.852***	.025		
I think this user has courage	.854***	.026		
I think this user is justice	.851***	.026		
Dignity 2 (DT2) Please let us know your opinion in terms of the following:			.931	.929
I think this user has superior social status	.870***	.020		
I think this user has high social standing	.899***	.017		
I think this user has high social status	.899***	.017		
I think this user is respected by others	.841***	.024		
Dignity 3 (DT3) Please let us know your opinion in terms of the following:			.838	.838
I think this user is free to choose their purpose	.733***	.044		
I think this user is free to express themselves as they chooses	.747***	.042		
I think this user is free to pursue themselves as they chooses	.751***	.041		
I think this user is free to realize their personal goals for development	.774***	.041		
Perceived trustworthiness (PT) Please tell us what you think about this user.			.861	.861
Trustworthy	.844***	.026		
Reliable	.809***	.030		
Honest	.810***	.030		

Note. ****p < .001*

Table 4. Variable summary statistics.

Variables	Mean	SD
(1) Social stigma (SS)	4.75	1.60
(2) Dignity 1 (DT1)	5.16	1.19
(3) Dignity 2 (DT2)	5.12	1.43
(4) Dignity 3 (DT3)	5.67	.94
(5) Perceived trustworthiness (PT)	5.59	1.37

(6) Doxing	.47	.5
(7) Doxing Type I	.33	.47
(8) Doxing Type II	.33	.47
(9) Doxing Type III	.35	.48

Table 5. Latent variable correlations.

V	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	**.85**							
(2)	.37*	**.85**						
(3)	.50*	.84*	**.88**					
(4)	.04	.47*	.37*	**.75**				
(5)	.19*	.77*	.75*	.42*	**.82**			
(6)	.19*	.17*	.15*	.12	.14			
(7)	-.06	.17*	.12	.07	.10	-.06		
(8)	.03	.07	.05	-.06	.10	.03	-.48*	
(9)	.03	-.23*	-.16*	-.02	-.19*	.03	-.51*	-.51*

Note. *V= Variables; Square Root of the Average Variance Extracted is in the diagonal; * p < .05.*

5.1.3. Main analysis. To understand whether doxing affects the social stigma and perceived dignity, we conducted a multivariate analysis of variance (MANOVA) with social stigma and the three types of dignity measurements as dependent variables and the doxing condition as the fixed effects. Hotelling's T^2 test reveals that doxing affects one or more of the dependent variables ($F = 2.65$, $p < .05$). The tests of between-subjects effects show that doxing affects social stigma and threats to perceived dignity. The results are shown in Table 6. Therefore, the research hypothesis H1 is supported.

Table 6. Between-subject effects under MANOVA.

Source	Dependent Variable	Sum of Squares	Df	Mean Square	F	Sig.
Doxing versus self-disclosure	Social stigma	17.25	1	17.25	6.98	.00
	Dignity 1	7.30	1	7.30	5.28	.02
	Dignity 2	7.83	1	7.83	3.92	.04
	Dignity 3	2.42	1	2.42	2.78	.10

To understand whether the different types of doxing affect the social stigma and threats to perceived dignity differently, we also conducted a second MANOVA that used social stigma and the three types of dignity as dependent variables and the second and third types of doxing as the fixed effects, treating the first types of doxing as the baseline condition. Hotelling's T^2 test reveals that the first and second types of doxing don't differ in affecting dependent variables ($F = .72$, $p > .05$). In contrast, the third type of doxing and the first type of doxing differ in affecting one or more of the dependent variables ($F = 3.14$, $p < .05$). The tests of between-subjects effects are shown in Table 7, providing more detailed results. Therefore, H2 is partially supported in that the third type of doxing can affect the social stigma and dignity differently than type I doxing, which can bring the same effects as the second type of doxing. The types of doxing Type I and Type II have been found to have the same effects on stigma and perceived dignity, which may be due to the reason that they are related to deanonymizing and targeting, which are less a concern of being associated with negative stereotypes than delegitimizing (i.e., releasing the immoral activities) the doxees, as suggested by Li and Whitworth (2022).

Table 7. Tests of between-subjects effects.

Source	Dependent Variable	Sum of Squares	Df	Mean Square	F	Sig.
Type II versus Type I doxing	Social stigma	1.16	1	1.16	.45	.50
	Dignity 1	.86	1	.86	.64	.43
	Dignity 2	.54	1	.54	.27	.60
	Dignity 3	.85	1	.85	.96	.33
Type III versus Type I doxing	Social stigma	1.07	1	1.07	.42	.98
	Dignity 1	13.54	1	13.54	10.04	.00
	Dignity 2	9.23	1	9.23	4.63	.03
	Dignity 3	.41	1	.41	.47	.50

5.2. Results of the second study

5.2.1. Study 2 improvements. Following Study 1, we improved our experimental design and the survey on several points. First, we merged the first and the second types of doxing because they didn't differ in affecting the dependent variables. Thus, the second experiment is a two (doxing versus self-disclosure) by two (doxing type I and II versus type III) experimental design. Second, we included a control condition where only the basic Facebook page is provided. Third, we changed the first attention check question, asking which animals are found on the moon, to the same as the second one, asking to click 'Disagree', because people may feel differently about what animals appear on the moon, making our attention check ratios low. Fourth, we changed the manipulation check questions to "The additional Facebook post I examined was (a) posted by another Facebook user; b) posted by themselves" and "The content of the additional Facebook post is about a) an activity attended; b) other information". Fifth, we further developed a new dignity construct related to social media (Dignity 4), adapting measures for inherent dignity proposed in Siegel (2012) because our focal phenomenon happens on social media. We want to make our main construct more relevant to the phenomenon we focused on. The new construct has four items: (1) I think this user is free to use social media as they choose; (2) I think this user is free to post on social media about any topic; (3) I think this user is free to express their best self on social media; (4) I think this user is free to pursue personal goals on social media. After improving the

experiments, we collected data from another 100 AMT users, and 96 of them completed our experiment.

5.2.2. Data description and various checks. The sample characteristics are similar to the first study. For the space consideration, we do not provide details that are available from the authors.

Out of the 96 respondents, 65.6% reported identifying as men and 34.4% reported identifying as women. Participants were predominantly college-educated or higher, were young to middle-aged (25-55), and supported the Democratic party. Furthermore, the pass rate for the two attention checks and two manipulation checks was 98.95% (95/96), 97.92% (94/96), 73.97% (54/73), and 76.71 % (56/73), respectively.

5.2.3. Measures, reliability, and validity. Using the whole sample (n=96), we measure the construct reliability by both composite reliability (CR) and alpha. The results are similar to Table 3, showing our constructs, including the newly developed one, are reliable and valid. For the space consideration, we do not provide details.

5.2.4. Main analysis. To understand whether doxing affects social stigma and threatens dignity, we conducted a MANOVA analysis by putting social stigma and the four types of dignity measurements into dependent variables and the doxing condition as the fixed effects. Hotelling's T^2 test reveals that doxing affects one or more of the dependent variables ($F = 2.67$, $p < .05$). The tests of between-subjects effects show that doxing affects the first two types of dignity ($F = 6.20$, $p < .05$; $F = 3.24$, $p < .05$).

To understand whether the different types of doxing affect the social stigma and threaten dignity differently, we also conducted a MANOVA analysis by putting social stigma and the four types of dignity measurements into dependent variables and the third type of doxing as the fixed effects. Hotelling's T^2 test reveals that the doxing types affect one or more of the dependent variables ($F = 2.70$, $p < .05$). The tests of between-subjects effects show that the doxing type only affects the first type of dignity ($F = 6.73$, $p < .05$).

We also conducted a covariance-based structural equation model (CB-SEM) to test the proposed research model. The results are shown in Figure 4.

From the CB-SEM results, we show that the perceived trustworthiness only mediates the impact of the doxing on dignity, where it partially mediates the first type of dignity and fully mediates the second type of dignity. Therefore, H3 is supported, and H4 is partially supported. And we can conclude that perceived trustworthiness only mediates the impact on dignity, but not the social stigma. The potential reason for this finding comes from two aspects. On the one hand, when users are perceived as trustworthy, they tend to be respected (Rotter, 1980), which makes the perceived dignity higher. However, when users are less trusted as being able to protect their personal information after being doxed, it is possible that their perceived stigma is not affected because stigma is a deviation in their social image from socially accepted characteristics or stereotypes (Goffman, 1963).

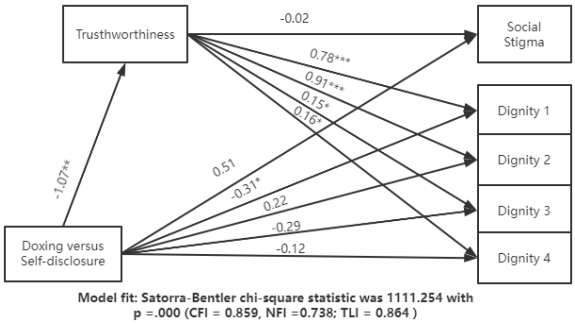

Figure 4. CB-SEM results.
Note. * p< .05; ** p < .01; *** p < .001

6. Discussion and conclusion

We proposed a research model that explains whether and how doxing affects the social stigma and dignity of users on social media platforms. Based on social identity threat theory and CARE theory, doxing, especially delegitimizing doxing, can produce discrimination and bring social stigma and threats to dignity. Using two online experiments, we test whether the doxing impact exists, whether different types of doxing matter, and why the main impact holds by explaining the underlying mechanisms. Doxing, especially the delegitimizing activity, may reduce the perceived trustworthiness of the doxees, which possibly leads to a lower level of dignity on social media platforms.

Across two studies, our results suggest that doxing affects social stigma and dignity, which addresses RQ1. Aware of the doxing impact, social media platforms should manage the doxing on this platform, such as using proactive or reactive strategies to mitigate the harms of the stigmatization and dignity of social media users. We also found that the doxing type matters that the doxing type III (i.e., delegitimizing) differs from other forms of doxing in how it affects social stigma and perceived dignity. This finding addresses the RQ2 about the moderation of doxing types on the doxing impact. Awareness of this impact can help social media platforms make customized strategies to mitigate the doxing impact. For example, social media platforms can treat the first type of doxing using the same strategy (e.g., reactive ones like removing the doxing content), while the third type should be carefully treated, such as making

policies to prevent (e.g., proactive strategies). Furthermore, by conducting a CB-SEM analysis, we show that perceived trustworthiness mediates the impact of doxing on perceived dignity, which addresses RQ3. Knowing this finding, social media platforms can intervene in the doxing impact by breaking the mechanism link of the perceived trustworthiness. For example, they can design a user trust system separating the impact from doxing.

6.1. Research contributions

Our study offers several contributions. First, we strengthen the understanding in the IS literature of doxing by empirically investigating its impact on social stigma and dignity using social identity threat theory and CARE theory. Our study responds to Fang et al. 's (2023) call for empirically examining doxing and its impact by focusing on understanding the mechanisms and doxing types on the impact on social stigma and perceived dignity.

Second, we extend the understanding of social stigma and perceived dignity on social media platforms. IS researchers have investigated human dignity (Leidner & Tona, 2021). However, little IS research has investigated the mechanisms that drive social stigma and adversely impact dignity. In this study, we examined doxing as a source that leads to social stigma and lowered perceived dignity. We found doxing types and perceived trustworthiness matter in explaining the doxing impacts.

Third, we extend the generalizability of social identity threat theory to the context of doxing on social media platforms. We use the social identity threat theory to explain that people being doxed can potentially be associated with negative images and stereotypes, making them stigmatized. We investigate the mechanisms and boundary conditions of the social identity threat theory by incorporating the mediating effect of perceived trustworthiness and the moderating effect of different types of doxing.

Practically, our research helps managers of social media platforms better understand the impact of doxing on social stigma and the perceived dignity of their users. Based on our research findings, the industry (e.g., Facebook) can better understand the consequences of doxing, especially its impact on the social stigma and users' dignity. Our work could help them make better policies regarding the doxing behaviors on their platforms.

6.2. Limitations and future research

This research has limitations that offer future research opportunities. First, we only studied the moderating roles of different doxing types on the main doxing impact. There may exist additional moderators, including doxing targets' social reputation (e.g., famous people versus ordinary people), doxing source credibility, and the anonymity of the doxers. Future work needs to investigate how doxees' characteristics moderate the impact of doxing as well. Also, doxing types could moderate the impact on perceived trustworthiness. Future research could also explore that.

Second, while we found support for the effect of perceived trustworthiness, additional mechanisms such as perceived overall similarity and suspicion of users that can help explain the main effects require further investigation. Third, we used self-disclosure as the control condition for the doxing condition, revealing the impact of the way of doxing. Future research could investigate the doxing versus non-doxing comparison and put the uniqueness of doxing, such as manipulating the malicious intention of releasing others' private information in the condition of doxing. Forth, the social stigma could also contribute to the perceived dignity. Hence, future research could explore whether social stigma may help explain the impact of doxing on perceived dignity. Lastly, in this research, we only consider the three types of doxing separately, where any combination of the three types of doxing can happen at the same time in reality. Future research is encouraged to investigate the impact of different combinations of the three types of doxing.

7. References

Banko, M., MacKeen, B., & Ray, L. (2020). A unified taxonomy of harmful content. In Proceedings of the fourth workshop on online abuse and harms, 125-137.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in Information Systems. MIS Quarterly, 35(4), 1017-1041.

Chan, T. K., Cheung, C. M., Benbasat, I., Xiao, B., & Lee, Z. W. (2022). Bystanders join in cyberbullying on social networking sites: the deindividuation and moral disengagement perspectives. Information Systems Research.

Chopra, K., & Arora, V. (2020) Covid-19 and social stigma: Role of scientific community. The Indian Journal of Tuberculosis, 67(3), 284-285.

Curtis, C. (2015). Anti-Social behaviour: A multi-national perspective of the everyday to the extreme. Anti-Social Behaviour, 1-120.

Douglas, D. M. (2016). Doxing: A conceptual analysis. Ethics and Information Technology, 18(3), 199-210.

Edlund, M., Lindwall, L., Post, I. V., & Lindström, U. Å. (2013). Concept determination of human dignity. Nursing Ethics, 20(8), 851-860.

Fang, Y., Risius, M., & Cheung, C. (2023). Understanding the current state of knowledge and future directions of doxing research: A social cognitive theory perspective. Hawaii International Conference on Information Systems.

Farnum, J., & Karimi, Y. (2020). Doxector: A web application for detecting doxing in Twitter.

Franz, A., & Thatcher, J. (2023). Doxing and doxees: A qualitative analysis of victim experiences. European Conference on Information Systems, Kristiansand, Norway.

Goffman, E. (1963). Stigma: notes on the management of spoiled identity, Simon and Schuster.

Griskevicius, V., Cialdini, R. B., & Kenrick, D. T. (2006). Peacocks, Picasso, and parental investment: The rffects of romantic motives on creativity. Journal of Personality and Social Psychology, 91(1), 63.

Hann, I. H., Hui, K. L., Lee, S. Y. T., & Png, I. P. (2007). Overcoming online information privacy concerns: An information-processing theory approach. Journal of management information systems, 24(2), 13-42.

Lee, J. K. (2015). Research framework for AIS grand vision of the bright ICT initiative. MIS Quarterly, 39(2), iii-xii.

Leidner, D. E., & Tona, O. (2021). The CARE theory of dignity amid personal data digitalization. MIS Quarterly, 45(1), 343-370.

Leidner, D. E., Tona, O., Wixom, B. H., & Someh, I. A. (2021). Putting dignity at the core of employee data use. MIT Sloan Management Review, 63(1), 1-7.

Li, Y. T., & Whitworth, K. (2022). Data as a weapon: The evolution of Hong Kong protesters' doxing strategies. Social Science Computer Review, 08944393221111240.

Li, Y. T., & Whitworth, K. (2023). Coordinating and doxing data: Hong Kong protesters' and government supporters' data strategies in the age of datafication. Social Movement Studies, 1-18.

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. Annual Review Sociology, 27(1), 363–85.

Martiny, S. E., & Nikitin, J. (2019). Social identity threat in interpersonal relationships: Activating negative stereotypes decreases social approach motivation. Journal of Experimental Psychology: Applied, 25(1), 117.

Mason, R. (2021). Policy for ethical digital services. Journal of the Association for Information Systems, 22(3), 11.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709-734.

McDonald, L. (2020). Shaming, blaming, and responsibility. Journal of Moral Philosophy, 1, 1-25.

Meese, H., Baker, T., & Sisson, A. (2020). # Wearebeneficiaries: Contesting poverty stigma through social media. Antipode, 52(4), 1152-1174.

Mitra, S., & Ransbotham, S. (2015). Information disclosure and the diffusion of information security attacks. Information Systems Research, 26(3), 565-584.

Mosquera, P. M. R., & Imada, T. (2013). Perceived social image and life satisfaction across cultures. Cognition & Emotion, 27(6), 1132-1141.

Öztürk, R., Bloom, T. L., Li, Y., & Bullock, L. F. (2021). Stress, stigma, violence experiences and social support of US infertile women. Journal of Reproductive and Infant Psychology, 39(2), 205-217.

Ritsher, J. B., Otilingam, P. G., & Grajales, M. (2003). Internalized stigma of mental illness: Psychometric properties of a new measure. Psychiatry Research, 121(1), 31-49.

Roth, P. L., Bobko, P., Shan, G., Ferrise, E., Roth, R., & Thatcher, J. (2023a). Doxing, political affiliation, and type of information: Effect on hiring related judgments. In Academy of Management Proceedings (Vol. 2023, No. 1, p. 10165). Briarcliff Manor, NY 10510: Academy of Management.

Roth, P. L., Bobko, P., Shan, G., Ferrise, E., Roth, R., & Thatcher, J. (2023b). Doxing, political affiliation, and type of information: Effect on suspicion, perceived similarity, and hiring related judgments. Journal of Applied Psychology.

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. American Psychologist, 35(1), 1.

Satorra, A., & Bentler, P. M. (2001). A scaled difference Chi-Square test statistic for moment structure analysis. Psychometrika, 66(4), 507-514.

Siegel, R. B. (2012). Dignity and sexuality: Claims on dignity in transnational debates over abortion and same-sex marriage. International Journal of Constitutional Law, 10(2), 355-379.

Snyder, P., Doerfler, P., Kanich, C., & McCoy, D. (2017). Fifteen minutes of unwanted fame: Detecting and characterizing doxing. Proceedings of the 2017 Internet Measurement Conference, 432-444.

Solo, A. M. (2019). Combating online defamation and doxing in the United States. Proceedings on the International Conference on Internet Computing (ICOMP): The Steering Committee of The World Congress in Computer Science, Computer, 75-77.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. in Advances in Experimental Social Psychology. Elsevier, 379-440.

Tan, A. (2022). To dox or not to dox, That is the question. Available at SSRN 4369643.

Trottier, D. (2017). Digital vigilantism as weaponisation of visibility. Philosophy & Technology, 30, 55-72.

West, B., Foster, M., Levin, A., Edmison, J., & Robibero, D. (2014). Cyberbullying at work: In search of effective guidance. Laws, 3(3), 598-617.

Yang, L. H., Kleinman, A., Link, B. G., Phelan, J. C., Lee, S., & Good, B. (2007). Culture and stigma: Adding moral experience to stigma theory. Social Science & Medicine, 64(7), 1524-1535.

Yudiana, T. C., Rosadi, S. D., & Priowirjanto, E. S. (2022). The urgency of doxing on social media regulation and the implementation of right to be forgotten on related content for the optimization of data privacy protection in Indonesia. Journal of Law, 9(1), 24-45.