Factors Influencing University Students' Online Disinhibition Behavior – The Moderating Effects of Deterrence and Social Identity

Wang, Kuei-Ing, Jou-Fan Shih

Abstract—This study adopts deterrence theory as well as social identities as moderators, and explores their moderating affects on online toxic disinhibition. Survey and Experimental methodologies are applied to test the research model and four hypotheses are developed in this study. The controllability of identity positively influenced the behavior of toxic disinhibition both in experimental and control groups while the fluidity of the identity did not have significant influences on online disinhibition. Punishment certainty, punishment severity as well as social identity negatively moderated the relation between the controllability of the identity and the toxic disinhibition. The result of this study shows that internet users hide their real identities when they behave inappropriately on internet, but once they acknowledge that the inappropriate behavior will be found and punished severely, the inappropriate behavior then will be weakened.

Keywords—Seductive properties of Internet, Online Disinhibition, Punishment Certainty, Punishment Severity, Social Identity.

I. INTRODUCTION

WITH the widespread use of technology, people rely on computers and electronic devices more and more. Online communication frameworks are showing rapid usage and growth over the last decade; computer bulletin boards, instant messaging, video chats and social networks are just a few examples of that. Such frameworks have become convenient for larger groups of internet users, who can connect with one another from their home computer, laptop or mobile smart phone. One phenomenon that has been shown to be characteristic of online communication participants is the online disinhibition effect, defined as a lowering of behavioral inhibitions in the online environment [1].

Computer-mediated communication (CMC) has since been found to have a correlation with uninhibited behaviors among users [4]. Many of the human behaviors displayed online (including violence, incitement, flaming, and verbal attacks, on the one hand, and self-disclosure, kindness, and the dispensing of help and advice, on the other) may be attributed to the online disinhibition effect.

In previous studies, it has been argued that behavior on the internet differs from similar behavior in the "real world" [2]. Such uninhibited or anti-normative online behaviors were sometimes interchangeable with other terms like “flaming” [3] and encompassed behaviors ranging from being impolite to using capital letters or exclamation marks and expressing personal feelings toward another person using a computer network [4]. Besides, Lapidot-Lefler and Barak pointed out that the negative online disinhibition effect is the concept used to refer to the negative results of this loss of inhibitions, usually manifested in aggressive behaviors that apparently would not be exhibited in a similar scenario in the “real world [5].” Suler called this anti-normal behavior on the internet—visiting places of pornography, crime, and violence, which are fields people might never explore in real-life—the “toxic disinhibition [6].”

This research tries to explore significant factors that explain why people feel uninhibited and behave in deviant ways online. Although there are many policies to avert illegal behavior in cyberspace, many people still think that they are safe behind the computer. Of all research studies centering on Internet attributes related to toxic disinhibition, Social Presence Theory Short and Reduced Social Cue Theory are often cited. Besides these two, the seductive properties of Internet are also a key feature when making prediction of Internet behavior [7]. Controllability, referred to the controlling power one feels he hold over the simulated world on the Internet, also results in less inhibited behaviors [8]. Fluidity of identity is yet another distinctive Internet attribute which, due to the anonymity and dissociative imagination linked with Internet use, creates and triggers anti-normative behavior [9]. Environmental dimensions are also very important factors. Containment theory assumes that every individual has an external containing structure as well as a protective internal structure. Both structures buffer, protect and insulate an individual against delinquency [10].

This study tries to find the significant factors that affect the students’ online toxic disinhibition. With the aim to pinpoint the key elements in determining toxic disinhibition, this research intends to accomplish the following objectives:

1) Integrating the roles of internet attributes and psychological factors, this study provides an innovative conceptual model in which social identity and deterrence are found and thus viewed to play important moderating roles, instead of being just another antecedents.
2) Building an integrated research model that synthesizes different aspects to explain the relationship between toxic disinhibition and Internet attributes, and that of Internet.
attributes and toxic disinhibition

II. LITERATURE REVIEW

A. Disinhibition

Lea [3] claimed that disinhibition includes flaming behaviors or hostile communication and has encompassed behaviors from being impolite and expressions of personal feelings toward another person using a computer network [4]. Toxic disinhibition describes the phenomena of online flaming and acting-out behaviors that often involve damaging another’s or even one’s own self-image, without any beneficial personal growth [3].

However, Joinson argued that if inhibition is when behavior is constrained or restrained through self-consciousness, anxiety about social situations, worries about public evaluation and so on [2], then disinhibition can be characterized by an absence or reversal of these same factors. In other words, disinhibition refers to the experience of a person whose behavior is no longer controlled by concerns about self-presentation or the judgments of others [11]. Others focus on the effect of disinhibition, as compared to behavior in the real world, since people may say or do different kinds of things online. People feel relaxed and express themselves much more openly, and their behavior is more uninhibited, called the “online disinhibition effect” [6].

In addition, the online disinhibition effect is a powerful mechanism that operates in interpersonal interactions in cyberspace, influencing people to act in ways they apparently would not have acted in their physical environment [12]. The anonymity of computer-mediated communication (CMC) often brings about less control in people so that they feel more uninhibited, and many kinds of rude language, critical opinions, anger, feelings of hostility, and even threats have been observed. People might also visit some places on the Internet, such as sites for pornography, crime, and violence: spheres they might never explore in real life. This is called toxic disinhibition [6].

B. Seductive Properties of the Internet

Suler proposed that some elements of the cyberspace result in online disinhibitions, such as dissociative anonymity, invisibility, asynchronicity, solipsistic projection, dissociative imagination, and minimalization of authority [6]. The relative anonymity and the lack of the usual “gating features” of Internet interactions greatly reduce the risks of such disclosure, and one can share intimate aspects of the self with much less fear of disapproval and sanction [13]. In addition, some other researches also argued that the Internet represents a place where they can exercise greater control over the impression that others form of them [4]. Turkle suggested two dimensions worthy of consideration: (1) the pleasure of control (i.e., the pleasure of being able to control the simulated world inside the computer such as in video games or online games), and (2) the perceived fluidity of identity in online life (i.e., the anonymous nature of chat rooms and ICQ, which allows participants to disguise their true identity in their interaction) [14].

Study showed that heavy users of the Internet are captivated by the ability to hold illusory power to control the world inside the computer (the pleasure of control), and by the capacity of the Internet to allow them to present a different persona (the perceived fluidity of identity in online life) [11], [15]. Leung found that “the perceived fluidity of identity in online life” significantly related to “social bonding” and “social identity.” These findings demonstrated the ability of the Internet to create an illusory relationship to present a different persona to form online relationships and to make people feel important and fashionable [11]. The seductive properties of the Internet allow users to redefine their identities and perceive illusionary power to control, as well as gives people a feeling of status and modernity, which may elevate their self-esteem [11].

Because of the anonymity and dissociative imagination provided by the Internet, people may feel that the imaginary characters they “created” exist in a different space, separated and apart from the demands and responsibilities of the real world [6]. Therefore, the seductive properties of the internet attract people to online social control and cause them to develop a preference for social interaction [11]. Therefore, based on the literature reviewed above, it was predicted that:

H1: Controllability of internet positively influences online toxic disinhibition
H2: Fluidity of identity positively influences online toxic disinhibition

C. Social Identity

Social identity theory is discussed early in 1974, against individuals based on self-identification (personal identity) theory developed, referring to the individual in a social group or groups have a certain degree of understanding or awareness of its existence, and this person has feeling that they are part of such a group or groups, will make this man's ideas with which organizations or groups linked to produce a certain extent. Network community is text-based that is the lack of social presence and social cues. Therefore, the anonymity online and the lack of visual cues, the new network provides the opportunity to people that they can develop a new relationship and social identity of the virtual community. Social identity of the online community with two common concepts, one is that people identify the community themselves, which is the feeling of belonging to that community; and the other concept is the identity of self-concept [16]. Although CMC gives us the opportunity to traverse social boundaries, paradoxically, it can also afford these boundaries greater power, especially when they define self- and group identity. Thus, anti-social behavior such as organizational norms have an advantage over the case, members of the organization will be very strong organizational norms compliance, antisocial behavior traction power will diminish. Organizations such as the use of social identity processes, to strengthen the organization's members agree.

Based on the above literature review, we propose the following hypotheses:

H3a: The magnitude of the Controllability of internet impact
on online disinhibition is associated with the level of Social Identity

H3a: The magnitude of the Fluidity of Identity of internet impact on online disinhibition is associated with the level of Social Identity

D. Deterrence Theory

Deterrence theorists believe that if punishment is severe, certain and swift, a rational person will measure the gains and losses before engaging in a crime and will be deterred from violating the law if the loss is greater than the gain [17]. Moreover, deterrence hypotheses can be tested by examining the relationship between people's current perceptions of the threat of sanctions and their current estimates of the probability of offending. Moreover, Klepper and Nagin presented respondents with scenarios involving tax noncompliance and systematically varied conditions likely to affect perceptions of the certainty and severity of punishment and the rewards of noncompliance [4]. Thus, respondents were contemplating the risks and rewards of offending within a specific context. Certainty of punishment simply means to make sure that punishment takes place whenever a criminal act is committed. If individuals know that their undesirable acts will be punished, they will refrain from offending in the future. Moreover, punishment must be swift in order to deter crime. The closer the application of punishment is to the commission of an offense, the greater the likelihood that offenders will realize that crime does not pay. Severity of punishment means that more severe a punishment, it is thought, the more likely that a rationally calculating human being will desist from criminal acts [23]. To prevent crime, therefore, criminal law must emphasize penalties. Punishment that is too severe is unjust, and punishment that is not severe enough will not deter criminals from committing crimes. According to containment theory [9], the determining factor that explains both conformity and deviance is the extent to which an individual is prohibited from committing delinquent acts. This prohibition, or control, comes from two sources: the self (inner containment) and the immediate social world within which the individual lives (outer containment) [18].

Based on the above literature review, we propose the following hypotheses:

H4a: The magnitude of the Controllability of internet impact on online disinhibition is associated with the level of Punishment certainty

H4b: The magnitude of the Controllability of internet impact on online disinhibition is associated with the level of Punishment Severity

H4c: The magnitude of the Fluidity of Identity impact on online disinhibition is associated with the level of Punishment Severity

H4d: The magnitude of the Fluidity of Identity impact on online disinhibition is associated with the level of Punishment Severity

III. RESEARCH MODEL AND METHODOLOGY

This study constructs a research model based on the seductive properties of the Internet that influence online disinhibition behavior. Also, we regard social identity, deterrence that moderates the relation between seductive properties of the Internet and online toxic disinhibition behavior. The research model is shown in Fig. 1.

A. Sampling

The target population of our study is university students. This research divided target population into two groups, one group (control group) filled out paper-based questionnaires directly, and the other group (experimental group) was given a video before filling out questionnaires. The content of video was about the punishment that would be given to who behaved antisocial behavior, such as use rude language, harsh criticisms. This research tended to explore if there any differences exist on different groups.

Questionnaires were used to collect data from university students. 568 questionnaires were received. On the cover of the questionnaire, we explained the goal of this research and gave some statements to ensure their privacy in filling up the questionnaire.

B. Measure Development

Measurement items were developed based on a comprehensive review of the literature as well as on expert opinion. A review of literature was undertaken to identify construct definitions and any existing measures. To the extent possible, previously published items were adopted or adapted. In short, we formed the scale for each construct in the model with developed and valid measures. Besides, this study adopts the Likert Scales, letting the participants choose from one to seven levels of agreement, with anchors ranging from 1 (strongly disagree) to 7 (strongly agree).

C. Demographic Analysis

561 valid samples were collected, in which 279 for experimental group and 282 for control group. The demographic information of these respondents is shown in Table I. The target population is university students, the data shows number of male and female, difference grade in university, as well as time spending online respondents respectively. More than half respondents spend more than 4 hours a day on surfing the Internet.
D. Measurement Model

This research adopts Partial Least Square (PLS) to assess the psychometric properties of the scales and to test the research model and hypotheses. PLS is a latent structural equation modeling technique that utilizes a component-based approach to estimation. According to Chin et al., PLS is perfectly suited for testing complicated relationships by avoiding inadmissible solutions and factor indeterminacy, and its capability in exploring complex relationships has been proven in many other studies [19].

This study conducted data analysis using structural equation modeling (SEM) implemented in partial least squares (PLS). While several methods can be used to analyze the data, we chose PLS. Because PLS is more appropriate when the research model is in an early stage of development and has not been tested extensively. A review of the literature suggests that empirical tests of affecting factor of the disinhibition are still sparse. This work is a preliminary effort to test in the standards literature. Hence, PLS is the appropriate technique for our research purpose. Thus, we utilized the Smart PLS software to analyze our study.

E. Constructs and Measurements

The literature review allowed us to identify construct definitions and any existing measures. In order to develop the scales for each construct in our model, we adopted and adapted previously developed, valid measures from published studies to fit our research context. The operational definitions are shown in Table II.

The pretest and pilot test are used to ensure that informants understood constructions, questions, and response scales of the study in the intended ways. Based on the respondents’ feedback, the questionnaire was adjusted to improve readability and accuracy.

F. Reliability and Validity

Reliability refers to the internal consistency of each construct. Composite reliability, Cronbach’s alpha and factor loading are employed to test reliability in this study. It has them approaching or exceeding the 0.7 level [21]. As shown Cronbach’s alpha should both be higher than 0.7, with most of been suggested that composite reliability (CR) and in Table II, all CR and Cronbach’s alpha results for each construct exceed 0.7, indicating adequate internal consistency of the measurements in our study.

Convergent validity refers to the extent to which items within a given construct should be related. Convergent validity should be validated when multiple indicators are used to measure one construct. It can be examined by item-total correlation (ITC), factor loading and average variance extracted by constructs (AVE) [20]. For adequate convergent validity, AVE should be greater than 0.5. The result in Table III shows proper convergent validity of the measurement items.

G. Data Analysis

Hypothesis testing was conducted through partial least squares regression analysis using smart PLS software. In order to examine if each hypothesis was supported, we assessed the t-statistic for the standardized path coefficients. In Table IV, the analysis is for 561 data from 2 groups. There is shown that hypothesis H1 was supported and H2 was not supported, it means controllability positively influences disinhibition behavior while fluidity of identity has no influences.

Social identity and punishment severity moderated the relation between Controllability and disinhibition, since the t values drop from 3.618 to 1.504 and from 3.618 to 0.826 in H1 respectively. Punishment severity didn’t moderate the relation between controllability and disinhibition.

In Table V, the analyses result is from experimental group. There is shown that hypothesis H1 was supported and H2 was not supported, it means Controllability positively influences disinhibition behavior while fluidity of identity has no Influences on. Social identity has moderated the relation between Controllability and disinhibition, since the t value from 1.932 to 0.64. Punishment certainty also has moderated the relation between Controllability and disinhibition (H4c: β=-0.537, t=1.909*).

In Table VI, the analyses result is from control group. There is shown that hypothesis H1 was supported and H2 was not supported, it means Controllability positively influences disinhibition behavior while fluidity of identity has no influences on. Social identity has moderated the relation between Controllability and disinhibition, since the t value from 1.932 to 0.64. Punishment severity also has moderated the relation between Controllability and disinhibition (H4c: β=-0.537, t=1.909*).
influences on disinhibition behavior. Punishment certainty and punishment severity have moderated the relation between controllability and disinhibition, since $H_1$'s t values drop to 1.334 and 1.252 from 5.588 respectively. Social identity has no moderating effect on the relation between online controllability and disinhibition.

### TABLE VI
ANALYSES RESULTS FROM ALL RESPONDENTS (N=561)

<table>
<thead>
<tr>
<th>no moderator</th>
<th>Moderated by social identity</th>
<th>Moderated by punishment certainty</th>
<th>Moderated by punishment severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$t$-value</td>
<td>$\beta$</td>
<td>$t$-value</td>
</tr>
<tr>
<td>$H_1$</td>
<td>0.3</td>
<td>3.618**</td>
<td>0.532</td>
</tr>
<tr>
<td>$H_2$</td>
<td>0.051</td>
<td>0.377</td>
<td>-0.05</td>
</tr>
<tr>
<td>$H_3a$</td>
<td>-0.265</td>
<td>666</td>
<td>0.129</td>
</tr>
<tr>
<td>$H_3b$</td>
<td>-0.021</td>
<td>0.054</td>
<td>-0.199</td>
</tr>
<tr>
<td>$H_4a$</td>
<td>-0.273</td>
<td>0.97</td>
<td>0.123</td>
</tr>
</tbody>
</table>

### TABLE V
ANALYSES RESULTS FROM EXPERIMENTAL GROUP (N=279)

<table>
<thead>
<tr>
<th>no moderator</th>
<th>Moderated by social identity</th>
<th>Moderated by punishment certainty</th>
<th>Moderated by punishment severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$t$-value</td>
<td>$\beta$</td>
<td>$t$-value</td>
</tr>
<tr>
<td>$H_1$</td>
<td>0.189</td>
<td>1.932**</td>
<td>0.276</td>
</tr>
<tr>
<td>$H_2$</td>
<td>0.13</td>
<td>1.488</td>
<td>0.576</td>
</tr>
<tr>
<td>$H_3a$</td>
<td>-0.111</td>
<td>0.223</td>
<td>-0.532</td>
</tr>
<tr>
<td>$H_3b$</td>
<td>-0.012</td>
<td>0.1</td>
<td>-0.155</td>
</tr>
<tr>
<td>$H_4a$</td>
<td>0.053</td>
<td>0.122</td>
<td>0.053</td>
</tr>
</tbody>
</table>

### TABLE VI
ANALYSES RESULTS FROM CONTROL GROUP (N=282)

<table>
<thead>
<tr>
<th>no moderator</th>
<th>Moderated by social identity</th>
<th>Moderated by punishment certainty</th>
<th>Moderated by punishment severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$t$-value</td>
<td>$\beta$</td>
<td>$t$-value</td>
</tr>
<tr>
<td>$H_1$</td>
<td>0.396</td>
<td>5.588*</td>
<td>0.567</td>
</tr>
<tr>
<td>$H_2$</td>
<td>-0.003</td>
<td>0.04</td>
<td>-0.245</td>
</tr>
<tr>
<td>$H_3a$</td>
<td>-0.201</td>
<td>0.651</td>
<td>0.334</td>
</tr>
<tr>
<td>$H_3b$</td>
<td>0.095</td>
<td>0.346</td>
<td>-0.0473</td>
</tr>
<tr>
<td>$H_4a$</td>
<td>0.092</td>
<td>0.392</td>
<td>0.055</td>
</tr>
</tbody>
</table>

### IV. DISCUSSION
This study introduced a comprehensive model to explore what factors lead to toxic disinhibition argued that the negative online disinhibition effect is the concept used to refer to the negative results of this loss of inhibitions, usually manifested in aggressive behaviors that apparently would not be exhibited in a similar scenario in the “real world.” We tested three dimensions (seductive Properties of the Internet, social and environmental) which lead to inhibit anti-normative behaviors online.

We first discuss the seductive Properties of the Internet, controllability and fluidity on toxic disinhibition. Controllability shows a significant positively effect on disinhibition, which is consistent with previous studies in that anonymity increases uninhibited behavior. Internet users can hide or alter their identity, or cause others to misidentify them, allowing themselves to perform aggressive behavior much more easily.

However, Fluidity of Identity has no significant impact on toxic disinhibition. This result is different with past studies that the reduced social cues (e.g., race, wealth, and status, which allow people to categorize others and organize information about them) facilitate a safe and fairly non-threatening form of social interaction in which individuals can self-disclose more, and do not lead to irregular or antisocial behavior. Here we attempt to give some...
explanations for this contradictory result. In some situations, people may inhibit their online behavior because they cannot see others online, and they worry that their aggressive words might hurt a person they actually know.

Next, we discuss the social dimension. The social identity has different effects on toxic disinhibition. In experimental group, social identity is negatively moderating the relation between controllability and toxic disinhibition. We imply that the identification of one’s belonging community will weaken the disinhibition behavior.

Last, we discuss the environmental dimension. Punishment severity has negative moderating effect on the relation between controllability and toxic disinhibition behavior in both experimental and control group. It means that students once realize that online disinhibition behavior will be punished severely that will weaken the intention to disinhibition behavior. However punishment certainty negatively moderated the relation between controllability and toxic disinhibition. It can be explained that if the online toxic disinhibition would be found certainly, students will hesitate to do some toxic disinhibition to avoid being caught.

V. CONCLUSION AND IMPLICATIONS

Drawing upon the integrated perspectives of the seductive properties of internet, social, and environmental dimensions, this study developed a conceptual model to examine influential factors in antisocial behavior online. We found that the controllability on internet has positive influences on toxic disinhibition on both experimental and control group. Furthermore, we found that fluidity of identity didn’t have moderating effect on disinhibition.

As for social identity has significantly moderated the relationship between controllability and toxic disinhibition, this study also proves that punishment certainty and severity negatively moderates the relation between controllability and toxic disinhibition. Meanwhile, punishment certainty and severity shows no moderating effects on the relation between fluidity of identity and toxic disinhibition. Below, we provide some suggestions for practitioners and elaborate on how the combined model can be used for further research.

Current internet upgrades rapidly in its transmission speed, the popularity of smart phones, information is transferred fast and plentiful by and from users, but it also brings a lot of negative effects, such as internet rumors, abuse and other misconduct. This research model can help researchers understand what factors will cause toxic disinhibition. When researchers do similar studies in the future, they can refer to the results of this research to further understand how to deal with the problem of toxic disinhibition online. In reviewing the literature, this study found that previous studies have addressed impact factors only separately, and that most of the prior studies merely concentrated on a single perspective of toxic disinhibition: either the internet dimension perspective or the psychological factor perspective.

Toxic disinhibition behavior is a kind of criminal behavior, however, there are few researches applying crime theory to examine toxic disinhibition behavior. This study found that deterrence has significant effect on toxic disinhibition. In practical, when education policy makers are formulating educational principles, the results of this study can call their attention not only to consider the internet characteristics but also consider the social and environmental dimensions. As a result, their policies will effectively help people avoid making inappropriate remarks and exhibiting inappropriate behavior online.

Since this study found that controllability would lead to disinhibition, managers of website can try to create environments that do not allow internet’s user the chance to generate these psychological factors. For instances, using student ID as an account can reduce online anonymity. In addition, we also found that the social identity as well as punishment certainty can lower the intention to toxic disinhibition, it is critical to have family, schools, and society participate in online ethical education programs. In school, teachers should emphasize the importance of moral issue. In society, government should strengthen public awareness to their citizen and promote the concept of moral principles via the mass media and company use ethic code to enhance their employee’s morality.

ACKNOWLEDGMENT

The authors would like to thank the National Science Council, Taiwan, R.O.C. for financially supporting this research under Contract No. < NSC 101-2511-S-159-001 >.

REFERENCES
