ISSN: 2688-8653

Anonymous Complicity: Research on the Criminal Psychological Mechanism and Criminal Law Imputation System of Incitement Behaviors in Cyberbullying

Xinyi Li

School of Civil and Commercial Court, Northwest University of Political Science and Law, Xi'an 710122, China

Abstract

With the popularization of online social platforms, cyberbullying incidents have shown an upward trend, among which the behaviors of inciters significantly promote the escalation of violence. Judicial practice faces challenges such as vague definition of subjective viciousness and difficulty in proving causation when identifying the liability of cyberbullying inciters. Inciters often have the characteristics of herd mentality and disinhibition, and the spread of their remarks will produce an obvious "snowball effect". The existing norms have overly general standards for identifying subjective intent in criminal law imputation, making it difficult to accurately distinguish between direct intent and indirect intent, leading to frequent sentencing imbalances. To improve the legislative system, it is suggested to construct an evaluation model for the degree of criminal psychological viciousness, incorporate factors such as criminal motivation and continuity of behavior into sentencing considerations, and learn from the red flag principle to strengthen platform review responsibilities. Establish quantitative standards for "serious circumstances" and clarify the determination boundary of inciting remarks through judicial interpretation, so as to effectively curb the spread of cyberbullying while protecting freedom of speech, and provide an operable theoretical basis for judicial organs to handle new types of cyberbullying cases.

Keywords

Cyberbullying Crimes; Inciters; Subjective Viciousness; Criminal Psychology; Criminal Law Imputation.

1. Introduction

With the proliferation of mobile internet and social media, cyberbullying has escalated into a significant social hazard. Inciters, distinct from ordinary emotional netizens, strategically exploit online anonymity and group psychology-using disinhibition and diffusion of responsibility-to set agendas, escalate conflicts, and manipulate public opinion. Their motives range from profit and retaliation to sheer malice, significantly lowering the threshold for initiating cyberbullying and amplifying its destructiveness, thereby creating a "few incite, many participate" harm model. However, China's current criminal legal framework lags behind this reality, struggling with vague subjective guilt assessment, difficult causation proof, and mismatched charges like insult or defamation, which fail to address the unique legal infringement of online incitement.

Therefore, a criminal psychology-based approach is urgently needed to analyze inciters' subjective mechanisms and construct a refined criminal imputation system. This includes developing an evaluative model for inciters' subjective malice-integrating motives, behavioral patterns, and speech content-to provide courts with a clear sentencing framework. Simultaneously, the feasibility of introducing a specific cyberbullying offense should be

ISSN: 2688-8653

examined, alongside refining standards for “serious circumstances,” applying the “red flag” principle to platform liability, and promoting a collaborative governance model combining legal regulation, technological empowerment, and platform responsibility. This approach also calls for re-examining theories like the “relativization of the duty of care” to rebalance risk distribution online, thereby reconciling free speech with the protection of personal rights.

2. Behavioral Characteristics and Regulation Status of Cyberbullying Inciters

2.1. Definition of Inciting Remarks

Compared with legitimate critical remarks, the essential difference of inciting remarks lies in their subjective intent of "triggering irrational group attacks"[1], rather than rational discussions based on facts. Critical remarks usually have specific direction and factual basis, aiming to promote problem-solving. Due to their networked nature, online inciting behaviors are more likely to trigger emotional resonance effects, making individual attacks escalate into group violence[2]. There is overlap but no equivalence between inciting remarks and insulting or defamatory behaviors in legal characterization. Insulting behaviors focus on the direct degradation of individual personality, and their harm is mainly reflected in the damage to specific victims. Defamatory behaviors take fabricating facts as the core element, while inciting remarks induce others to commit violence by creating reasons for attacks, and their harm is diffusive and secondary communicative. Although Article 246 of the Criminal Law stipulates insult and defamation side by side, inciting remarks often have both distortion of facts (defamation elements) and obvious personality degradation (insult elements). This compound illegal form makes it difficult for traditional charges to fully evaluate their social harm[3].

2.2. Behavioral Mode and Core Role of Inciters

Cyberbullying inciters play a crucial role in the development of incidents, and their behavioral modes usually show obvious phased characteristics. In the initiation stage, inciters trigger group attacks by creating topics and setting emotional trigger points, label specific individuals or groups with negative tags, and create opposition by exploiting the empathy weaknesses or moral anxieties of the public. Studies have shown that cyberbullying often starts with individual initiators making insulting, defamatory and other remarks against specific individuals, and then, with the rapid spread of the Internet, touches the universal demands of the public in terms of emotional catharsis and personal safety, thereby gaining widespread attention and participation[4].

After the wave of remarks is set off, inciters enter the amplification stage, realizing the amplification of sound waves by guiding the direction of public opinion and strengthening the intensity of attacks. They distort the overall picture of facts by selectively presenting information fragments, strengthen group identity through binary opposition discourse of "us vs. them", and deepen the public's stereotypes by repeatedly exposing the victim's "crimes". At this stage, inciters will fully utilize the disinhibition effect of the anonymous environment, and reduce the moral constraints of participants through emotionally appealing expressions. Some professional inciters even establish attack discourse databases and adopt standardized incitement templates for different types of incidents, which greatly improves the organizational level of violent behaviors. The concealment of online inciting behaviors significantly reduces the cost of crime, and inciters can easily hide their behavioral traces[5], further encouraging their courage to amplify attacks.

In the continuous stage, inciters are committed to maintaining the heat of attacks and preventing the calm of public opinion. They continuously dig out the victim's historical remarks for secondary attacks, organize punch-in attacks to create continuous pressure, block the

ISSN: 2688-8653

victim's voice channels through report bombing and other means, and fully utilize the snowball effect - when the scale of attacks reaches a critical point, even if active incitement is stopped, group behavior will still maintain itself, reflecting obvious subjective viciousness. The degree of harm is far greater than that of a single insult, but the existing law has not fully considered the cumulative harm over time in the identification of "serious circumstances". Through phased psychological manipulation, inciters transform loose netizens into digital mobs with attack directions, triggering group violence far beyond their direct influence. The current law has obvious lag in regulating such structured behaviors, especially the lack of effective response strategies to new means such as cross-platform collaborative incitement and periodic hot topic speculation, which urgently need to be solved through legislative improvement and judicial innovation.

3. Analysis of Criminal Psychological Mechanism of Inciters' Subjective Viciousness

3.1. Deindividuation and Responsibility Diffusion Psychology

The online anonymous environment provides a unique psychological protection barrier for inciters, making it easier for them to break through the moral constraints in real society. When individuals are in an anonymous state, they will produce an obvious "deindividuation" effect, that is, weakened self-awareness and increased compliance with group norms, making it easier for inciters to engage in aggressive behaviors online that they would not do in daily life[8]. The group environment strengthens the psychology of responsibility diffusion, leading participants to form the wrong cognition that "the law does not punish the majority".

The anonymous group environment leads to the systematic failure of moral constraint mechanisms. In real society, individual behaviors are subject to multiple constraints such as social evaluation and interpersonal relationships, while the virtuality of the online space cuts off this feedback mechanism. Taking advantage of this feature, inciters accelerate the process of moral desensitization by designing emotionally appealing symbols. Cyberbullying has the bullying nature of creating group mental torture, which is directly related to the failure of moral constraints[9]. Cognitive dissonance in deindividuation makes it difficult for individuals to connect online behaviors with their real selves, responsibility diffusion reduces the prediction of behavioral consequences, and the failure of moral constraints lifts the internal behavioral ban. The interaction of these three forms a psychological vicious circle - the more involved in group attacks, the more blurred self-cognition becomes, and the blurred self-cognition further prompts individuals to seek a sense of existence in aggressive behaviors. This psychological mechanism is an important reason for the dilemma of "the law does not punish the majority" in cyberbullying[10].

3.2. Psychological Dynamics in the Group Polarization Effect

The group polarization effect presents complex psychological dynamics in cyberbullying incidents, among which the behavioral motives of inciters and group interaction form a two-way reinforcement mechanism. Inciters often expand their influence by deliberately catering to group emotions. They are well aware of the communication law of emotional priority, package attack intentions with discourse such as moral condemnation, and essentially act as "emotional intermediaries" to manipulate group cognition by screening and amplifying specific information. Cyberbullying is a premeditated intentional behavior. Inciters form psychological coercion on victims through high-frequency and inescapable information bombing[11]. This strategic operation is essentially different from simple speech misconduct, and inciters themselves may also fall into the irrational group atmosphere they created. The violence of cyberbullying is not only reflected in direct harm, but also in the mental oppression formed

ISSN: 2688-8653

through verbal information[12]. At this time, inciters not only play the role of emotional manipulators, but also become victims of group polarization. Their subjective state gradually evolves from the initial utilitarian motivation to irrational emotional catharsis. The key to determining direct intent lies in identifying whether the inciter has a subjective pursuit of the occurrence of harmful consequences. In the identification of indirect intent, it is necessary to examine the inciter's "degree of indifference" to harmful consequences. The group polarization environment makes intent identification face the problem of "responsibility dilution". When multiple inciters collude in the crime, the boundary of individual subjective viciousness becomes blurred. In judicial practice, the "dominance theory" can be learned from, focusing on examining who created the unacceptable source of danger.

4. Difficulties in Criminal Law Identification of Inciters' Subjective Guilt and Causation

The significant difficulty in identifying the subjective intent of cyberbullying inciters stems from the particularity of the online environment and the complexity of the actor's psychological state[6]. Direct intent requires the actor to hope for and pursue the occurrence of harmful consequences, while indirect intent is manifested as the attitude of knowing that harmful consequences may occur but allowing them to happen[13]. In the online environment, this distinction often becomes blurred due to the complex psychological state of the actor.

The primary problem faced in judicial practice is that inciters often express their true intentions through obscure language. Online language has a high degree of context dependence and ambiguity, making it difficult for judicial organs to accurately judge their true psychological state. The identification of direct intent requires proving that the inciter has a clear target orientation, but in more cases, inciters will adopt indirect expressions and use potentially inciting remarks. At this time, their subjective state is more in line with the characteristics of indirect intent of "knowing that harmful consequences may occur but allowing them to happen"[14]. This ambiguity brings substantial difficulties to judicial identification, and it is urgent to establish intent identification standards suitable for the online context.

Some inciters will show contradictory psychology towards harmful consequences. On the one hand, they hope to expand the influence of their remarks, and on the other hand, they are skeptical or negative about the actual harmful consequences. Some scholars have proposed that the initiation and organization of cyberbullying are punishable in themselves, but this view has not yet formed a consensus in judicial practice[10]. The fragmented nature of online remarks makes it extremely difficult to restore the actor's complete psychological process. The correlation analysis of remarks on different platforms requires a lot of judicial resources, and the short data retention period of some platforms also leads to the easy loss of key evidence. A diversified intent identification system should be constructed. For the proof of direct intent, emphasis should be placed on factors such as the degree of organization and planning of the actor and the selectivity of targets; for indirect intent, a reasonable person standard should be established to judge the predictability of harmful consequences in combination with the cognitive level of ordinary netizens. Judicial interpretation should clarify the proof standard of indifferent psychology, explore the establishment of an online behavior trajectory analysis system, and restore the actor's psychological process through big data technology to provide objective basis for judicial identification.

ISSN: 2688-8653

5. Improvement of the Imputation Path for Online Inciting Behaviors of Cyberbullying

5.1. Development of Legislation-Oriented and Preventive Regulation

It is fully necessary and feasible to add the crime of online incitement to cyberbullying, and its constitutive elements should be carefully designed. Objectively, it requires the actor to publicly issue inciting remarks that are sufficient to induce an unspecified majority of people to carry out cyberbullying against specific targets. The theoretical concept of "independent dangerous crimes" can be introduced, that is, as long as the act creates a danger not tolerated by law, it is punishable[11]. The identification of "sufficient to induce" should establish a multi-level judgment standard system, including quantitative indicators and qualitative evaluation dimensions. Quantitative indicators can refer to objective data such as the dissemination scope, duration, and number of reposts of remarks. Qualitative evaluation needs to examine factors such as the particularity of the target of attack, the extreme degree of labels used, and the intensity of incitement of remarks.

When determining liability, a distinction should be made between direct and indirect intent: the former applies to core inciters who deliberately manipulate group polarization, while the latter applies to those who post inflammatory content for attention or emotional release knowing it may incite cyberbullying. Penalties should combine special and general prevention, supplementing imprisonment with corrective measures like online gag orders, community service, and psychological correction mandates, alongside fines for profit-driven incitement. Before establishing a specific charge, judicial interpretation should optimize existing statutes by redefining "serious circumstances" under Article 246 of the Criminal Law to include factors such as the act's inherent danger, degree of organization, and technical harm. Continuous psychological harm should independently constitute a serious offense, recognizing psychological causation between incitement and the victim's mental suffering[7]. A reversed burden of proof could be explored, requiring suspects to demonstrate the absence of social harm, thereby addressing online evidence challenges. Finally, a stepped legal liability system should be built to seamlessly connect administrative and criminal penalties, ensuring comprehensive legal oversight.

5.2. Distinction and Selection of Judicial Imputation Paths

In judicial practice, a hierarchical and clear imputation path selection mechanism should be constructed, and an appropriate criminal liability identification model should be selected according to the actual status, mode of action, and degree of harm of the actor in the cyberbullying incident. This distinction is related to the accuracy of conviction, directly affects the appropriateness of sentencing, and is an important guarantee for realizing judicial justice. The complexity of cyberbullying cases requires judicial organs to adopt more refined identification methods to accurately crack down on crimes, avoid excessively expanding the scope of crackdowns, and ensure that the application of penalties complies with the principle of proportionality. In judicial practice, attention should be paid to distinguishing the degree of subjective viciousness and behavioral harm of different actors, and establishing differentiated criminal liability identification standards.

A gradient standard for identifying platform liability should be established. Large-scale online platforms should be assigned higher duties of care and review responsibilities because they have stronger technical capabilities and richer management experience. For small and medium-sized platforms, the scope of review obligations should be determined according to their actual technical capabilities and operating scale. A good faith defense mechanism for platform liability exemption should be established, and liability exemption should be given to platforms that have taken reasonable measures to balance the interests of all parties. It is suggested to explore the

ISSN: 2688-8653

establishment of a stepped liability identification system, and set different liability identification standards according to the participation degree and role of the actor in the cyberbullying incident. For ringleaders such as organizers and planners, strict liability identification standards should be applied, and their criminal liability should be severely pursued; for active participants, relatively strict liability standards can be applied according to their specific behaviors and roles; for general participants, lighter liability identification standards or exemption from criminal liability can be considered.

6. Conclusion

Cyberbullying incitement, leveraging online anonymity and group dynamics, has become a complex criminal act that systematically manipulates public opinion and violates personal dignity. Inciters blend rational calculation with emotional indifference, using social psychology to morph scattered emotions into organized digital violence. Traditional criminal law, designed for physical space, struggles to address this due to three core dilemmas: proving intent amid ambiguous online language, establishing causation from countless aggregated anonymous acts, and applying mismatched traditional charges.

To overcome this, a three-part legal reform is proposed. First, legislation must shift from a "result-oriented" to an "act-oriented" model, creating a new "Incitement to Cyberbullying" offense focused on criminalizing the creation of legally defined dangers. Second, judiciary must refine attribution, distinguishing between "abettor" and "indirect perpetrator" roles based on an actor's actual influence for precise accountability. Third, sentencing should integrate scientific psychological assessments of an offender's moral cognition and recidivism risk to move beyond punishment based solely on objective harm.This framework aims to provide actionable judicial solutions. Future research should develop strategies for diverse online communities and leverage AI to predict and prevent inciting behavior, fostering a safer digital environment.

References

[1] Shi Jinghai. On the Substance of Cyberbullying and the Improvement of Criminal Law Application Rules[J]. Science of Law (Journal of Northwest University of Political Science and Law), 2023, (5):69-82.

[2] Xiao Bojin. Research on the Legal Regulation of Cyberbullying in China[D]. Master's Thesis of Shanxi University of Finance and Economics, 2024.

[3] Mei Chuanqiang. Research on the Generative Mechanism of Criminal Psychology[D]. Doctoral Thesis of Southwest University of Political Science and Law, 2004.

[4] Liu Pujun, Zhang Guihong. The Sanction Dilemma and Response Ideas of Online Incitement Crimes[J]. Social Sciences in China, 2017, (3):52-57.

[5] Guo Hongwei, Qu Xiaomeng. The Judicial Path of Cyberbullying Governance[J]. Journal of Shandong Judges Training Institute, 2025, (3):139-153.

[6] Supreme People's Court, Supreme People's Procuratorate, Ministry of Public Security. Guiding Opinions on Punishing Illegal and Criminal Acts of Cyberbullying in Accordance with the Law (Draft for Comment)[J]. People's Procuratorate Daily, 2023.

[7] Zhou Libo. The Criminal Law Governance of Cyberbullying Crimes[J]. Research on Rule of Law, 2023, (5):38-51.

[8] Tian Hongjie. The Dilemma of "The Law Does Not Punish the Majority" in the Criminal Law Governance of Cyberbullying and Its Resolution[J]. Law Science Magazine, 2024, (1):92-103.

[9] Xu Ying. On the Criminal Liability for Suicide and Death Caused by "Cyberbullying"[J]. Tribune of Political Science and Law, 2020, (1):133-142.

ISSN: 2688-8653

[10] Li Bencan. The Dogmatic Construction of the Punishment Scope of Cyberbullying Crimes[J]. Global Law Review, 2025, (1):5-12.

[11] Wang Huawei. A Systematic Evaluation and Reflection on China's Cybercrime Legislation[J]. Law Science Magazine, 2019, (10):82-93.

[12] Zhang Mingkai. A Probe into the Controversial Issues of Online Defamation[J]. China Legal Science, 2015, (3):60-79.

[13] Xu Caiqi. On the Criminal Law Regulation of "Cyberbullying" Behaviors[J]. Judicial Application, 2016, (3):102-108.

[14] Cai Rong. The Legitimacy and Dogmatic Analysis of Criminalizing "Online Verbal Violence"[J]. Journal of Southwest University of Political Science and Law, 2018, (2):63-72.