ISSN: 2688-8653

Adolescent "Self-Harm Games": Triadic Reciprocal Mechanisms and Preventive Strategies

Hao Miao[1], Yi Zhang[2,*]

[1] School of Education and Psychological Science, Sichuan University of Science and Engineering, Zigong 643000, Sichuan, China

[2] School of Education and Psychological Science, Sichuan University of Science and Engineering, Zigong 643000, Sichuan, China

[*]Corresponding Author

Abstract

"Self-harm games," an aberrant manifestation of non-suicidal self-injury (NSSI), have garnered significant societal concern due to their programmatic, covert, and coercive nature, which profoundly compromises the physical and mental well-being of adolescents. This phenomenon extends beyond simple behavioral imitation; rather, it emerges from a vicious interaction among environmental induction, cognitive reconstruction, and behavioral reinforcement. Protecting adolescents from the detrimental impacts of toxic subcultures and ensuring a healthy developmental environment is a critical public health priority. Consequently, it is imperative to establish a collaborative, multisectoral mechanism. By implementing multidimensional interventions-encompassing environmental regulation, cognitive guidance, and behavioral correction-the generative chain of self-harm games can be effectively severed, thereby safeguarding the healthy development of adolescents.

Keywords

Triadic Reciprocal Determinism; Adolescents; Self-Harm Games.

1. Introduction

In recent years, alongside the rapid development of the Internet and social media, adolescent non-suicidal self-injury (NSSI) has exhibited alarming new trends characterized by "online migration" and "gamification." This aberrant manifestation, termed "self-harm games," is not a traditional recreational activity; rather, it camouflages self-harming behaviors under the guise of "task challenges" and "social check-ins." Exploiting adolescents' psychological vulnerabilities-specifically their need for thrill-seeking and a sense of belonging-these games utilize mechanics such as points, leveling up, and rankings to reframe physical pain into a currency for acquiring group identity. From the early "Blue Whale" challenge to recent variants lurking in hidden online communities (e.g., "Dreaming of the Tang Dynasty" and "Alien Monkeys"), this behavior has transcended the traditional scope of individual emotional regulation. Instead, it has mutated into a collective behavior defined by highly structured rules, strict hierarchies, and distinct subcultural identities. Despite repeated targeted crackdowns by educational authorities and cyber-regulatory agencies, these self-harm games frequently resurge after periods of dormancy by adopting new aliases and evolving their argot, increasingly targeting younger demographics while demonstrating enhanced concealment.

ISSN: 2688-8653

2. Literature Review

The academic community has accumulated substantial research on adolescent self-harm. Early studies predominantly adopted a psychopathological perspective, focusing on individual depressive moods, impulse control disorders, and family dysfunction, thereby conceptualizing self-harm as a maladaptive coping mechanism for psychological distress. However, as research has deepened, scholars have increasingly recognized the critical role of environmental factors. Sherman et al. (2016) noted that peer feedback on social media can significantly activate the reward circuits in adolescents' brains, producing a potent behavioral reinforcement effect[5]. Khasawneh et al. (2020), through an analysis of YouTube and Twitter, further confirmed that extreme self-harm behaviors exhibit a marked "contagion effect" in cyberspace, where algorithmic recommendation mechanisms inadvertently construct a highly inductive information environment[6]. Furthermore, a recent study by Mancone et al. (2025) revealed the driving roles of boredom and peer pressure in hazardous game participation, indicating that contemporary self-harm behaviors are gradually alienating into a social endeavor aimed at identity-seeking[7].

Although existing studies have explored this phenomenon from multiple dimensions-including psychological motivations, online dissemination, and peer influence-they largely remain fragmented. There is a conspicuous lack of an integrated theoretical framework to explicate how the dynamic interaction among the "environment, cognition, and behavior" facilitates the acquisition and consolidation of self-harm games. Why do cognitively immature adolescents willingly jeopardize their physical and mental health to participate in these extreme games? How do online algorithmic environments "collude" with individual cognitive biases? Addressing these gaps, this study introduces Albert Bandura's Social Learning Theory, specifically its core tenet of "Triadic Reciprocal Determinism." By transcending single-factor attributions, this research constructs an integrated model encompassing environmental induction, cognitive reconstruction, and behavioral reinforcement. Ultimately, this study aims to delve into the generative logic of adolescent "self-harm games" and propose systematic, evidence-based prevention strategies.

3. Phenomenological Characteristics

The clandestine proliferation of "self-harm games" among adolescents and their subsequent adverse social consequences stem from the malicious appropriation and alienation of traditional gaming mechanics. Distinct from typical non-suicidal self-injury (NSSI), these behaviors exhibit pronounced programmatic, organizational, and coercive characteristics.

3.1. Programmatic Mechanisms: Progressive Manipulation Traps

The most prominent feature of "self-harm games" lies in their rigorous, programmatic game logic, which induces adolescents to sink deeper step by step through carefully designed task gradients.

These games typically establish highly deceptive entry requirements. Through an analysis of the "Blue Whale" death game, Lv (2020) observed that such games usually encompass thirty to fifty task directives of escalating difficulty [1]. Initial tasks often appear innocuous and carry a certain mysterious, ritualistic quality, such as "waking up at 4:20 AM," "listening to specific depressive music for an entire day," or "watching a horror movie." These low-threshold preliminary tasks exploit the psychological "foot-in-the-door effect": by first presenting a trivial request, perpetrators cultivate a compliant mindset within the individual before progressively introducing more demanding requirements. For adolescents yearning for thrill-seeking and uniqueness, this mysterious and easily accessible "admission ticket" frequently holds immense allure.

ISSN: 2688-8653

Essentially, the entire trajectory of "self-harm games" constitutes a malicious reverse application of systematic desensitization. As the game progresses, task directives shift from psychological disturbance to physical harm, such as "pricking a finger with a needle" or "carving specific totems on the arm." Research by Mancone et al. (2025) indicates that this gamified mechanism redefines physical harm as a requisite means of "leveling up" [7]. Through continuous incremental probing and repetition, adolescents' physiological fear of pain, bleeding, and even death is gradually eroded. The completion of each task is imbued with the false significance of "conquering oneself," causing adolescents' pain thresholds and psychological defenses to crumble imperceptibly. Ultimately, this psychological conditioning enables them to execute extreme suicidal directives in the final stages without experiencing fear.

3.2. Organizational Concealment: Disinhibited Community Environments

Distinct from public online social environments, the dissemination of "self-harm games" relies on highly concealed anonymous communities. This specific cyber environment acts as a massive catalyst for extreme behaviors.

On the one hand, the construction of anonymous communities for "self-harm games" possesses high anti-reconnaissance capabilities. Organizers frequently utilize platforms with self-destructing messages or encrypted communication features (such as Telegram and QQ) to establish contact, forming "dark web" islands isolated from conventional surveillance. Within these groups, members exclusively use aliases or nicknames, existing in a state of complete deindividuation.

On the other hand, the anonymous environment induces a profound disinhibition effect. Chen (2025), in a study on anonymous groups such as "Momo," found that when individuals' true identities are concealed, their sense of social responsibility significantly diminishes; not only are they more likely to evade real-world moral constraints, but they also tend to exhibit more radical aggressive behaviors than in their real lives [2]. In the specific "echo chamber" of self-harm game groups, the life values of mainstream society are thoroughly subverted. Self-harm behaviors are no longer viewed as pathological; rather, they are alienated into symbols of "courage," "loyalty," and "conformity."[3] Research by Khasawneh et al. (2020) also confirms that frequent interactions within these closed communities generate significant emotional contagion. Members mutually reinforce each other through practices like sharing images of self-injury and engaging in competitive suffering, causing individual rationality to be rapidly swallowed by the collective frenzy [5].

4. Theoretical Framework and Analytical Approach

4.1. Analytical Approach: Conceptual Analysis and Literature Synthesis.

As a theoretical and conceptual exploration rather than an empirical quantitative study, this research adopts a rigorous conceptual analysis and literature synthesis as its primary analytical approach. By systematically reviewing and critically synthesizing recent multidisciplinary literature-encompassing cyber psychology, adolescent development, and social media behavioral studies-this paper examines the underlying mechanisms of extreme online phenomena such as the "Blue Whale" and other self-harm variants. This approach allows for a comprehensive deconstruction of the socio-technical environments, specifically algorithmic echo chambers and anonymous communities, avoiding the limitations of traditional, single-factor psychopathological analyses. By abstracting the behavioral patterns of these digital subcultures, this study establishes a robust theoretical foundation for exploring the complex psychological and environmental dynamics driving adolescent participation in self-harm games.

ISSN: 2688-8653

4.2. Theoretical Framework: Triadic Reciprocal Determinism.

To elucidate the complex generative logic of these games, this study introduces Albert Bandura's Social Learning Theory, specifically anchoring on its core tenet: Triadic Reciprocal Determinism. According to this framework, human functioning is explained in terms of a continuous reciprocal interaction among cognitive, behavioral, and environmental determinants.Applied to the context of this study, the acquisition of "self-harm games" among adolescents is not a unidirectional process of imitation; rather, it is the product of a vicious, sustained interaction among environmental factors, cognitive states, and behavioral actions.This theory perfectly contextualizes the phenomenon: the algorithmic and anonymous network acts as the environmental induction; the symbolic translation of pain into game rewards drives cognitive reconstruction; and the peer-driven or fear-based feedback loops consolidate behavioral reinforcement.Based on this theoretical foundation, an integrated acquisition model of adolescent "self-harm games" is constructed to visualize these dynamic interactions (see Figure 1).

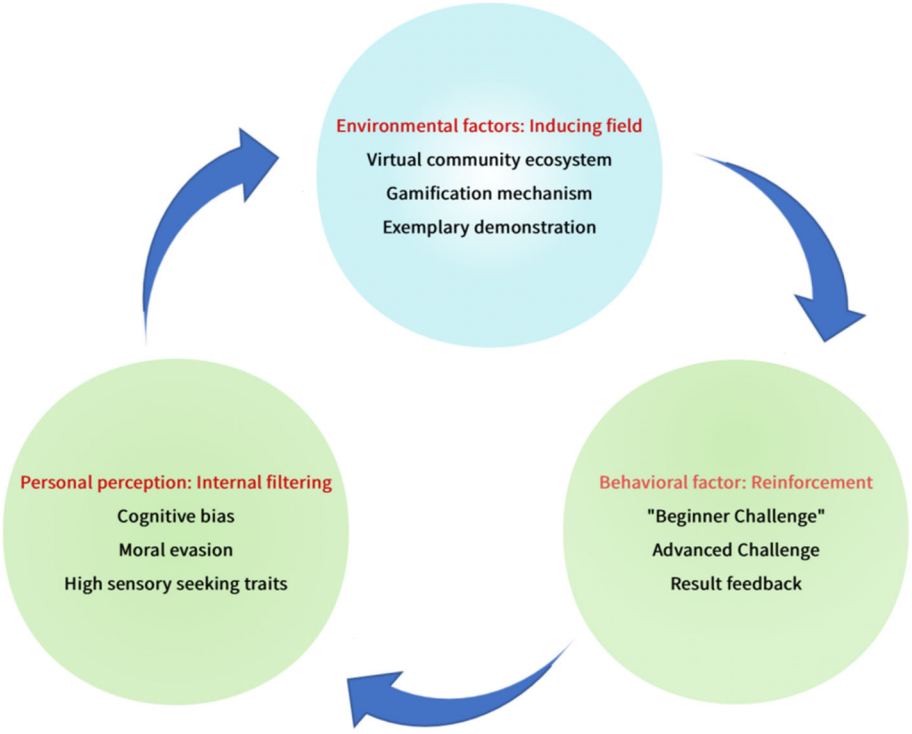

Figure 1. The behavioral acquisition model of adolescent "self-harm games" based on the triadic reciprocal perspective of social learning theory

5. Triadic Reciprocal Mechanisms

Drawing upon Bandura's Triadic Reciprocal Determinism, the acquisition of "self-harm games" among adolescents is not a unidirectional process of imitation; rather, it is the product of a sustained, vicious interaction among the environment, cognition, and behavior [8]. Consequently, this study constructs a generative model comprising "environmental induction-cognitive reconstruction-behavioral reinforcement" to analyze how such behaviors are acquired and consolidated within adolescent cohorts via Internet algorithms and gamification mechanisms.

5.1. Attention and Observation Stage

In the triadic reciprocal model, environmental factors induce the acquisition of individual behaviors. The cyber environment precisely captures the attention of adolescents through the

ISSN: 2688-8653

collusion of algorithms and deindividuation mechanisms. Specifically, algorithmic recommendation systems on social media construct a closed "information cocoon" exclusive to adolescents. Research by Khasawneh et al. (2020) indicates that algorithms exploit adolescents' sensitivity to novelty [5]. Once the system captures their browsing footprints, it continuously pushes similar extreme videos, subsequently generating high-density visual impacts. This precise content delivery traps adolescents in an "echo chamber effect". Rational external voices are filtered out by the system, allowing the closed environment to constantly rationalize and even glorify extreme self-harm behaviors. This environmental distortion reshapes adolescents' judgment, leading them to falsely believe that self-harm is ubiquitous and confers a sense of superiority, thereby significantly reducing their vigilance against potential dangers. Furthermore, anonymous online communities severely undermine social control. Chen (2025) noted that anonymity enables individuals to evade real-world moral constraints, making them more prone to disinhibited behaviors under the illusion of the "diffusion of responsibility" [2]. Consequently, individuals easily succumb to the misconception of impunity in numbers, thereby exhibiting unrestrained behaviors. In those private groups disguised as "games," engaging in self-harm becomes a mandatory threshold for entry. This subcultural atmosphere provides an ideal arena for imitation, where the environment subtly induces adolescents to direct aggression toward themselves.

5.2. Retention and Reproduction Stage

The manner in which individuals encode information directly determines whether a behavior can be persistently retained. The fundamental reason "self-harm games" can breach physiological pain defenses lies in their induction of cognitive translation among adolescents. Symbolic encoding thoroughly reconstructs adolescents' comprehension of pain. Bandura postulated that observers must transform behaviors into symbolic representations to retain them in memory. Within the gaming context, self-harm behaviors-originally representing "danger" and "pain"-are translated by rules into "points" or "levels" [8].Consequently, the brain's evaluation system undergoes substitution; negative experiences of harm are redefined as positive achievements. Semantic desensitization plays a critical role in the emergence of adolescents' "self-harm behaviors". When reproducing the actions, adolescents no longer focus their attention on physical agony, but rather on the precise execution of game rules; the warning signals issued by biological instincts are completely shielded at the cognitive level. Simultaneously, behavioral reproduction evolves into a form of "performative dissociation". Because games mandate participants to upload video evidence, self-harm is no longer confined to mere self-attack, but is transformed into a camera-facing "bodily performance". Research by Mancone et al. (2025)demonstrates that the process of examining one's own wounds through a screen highly precipitates "self-objectification" [7].The body is alienated into a "prop" or "canvas" for completing tasks. A psychological separation occurs between the observing subject and the injured object; this state of dissociation significantly weakens the psychological resistance when inflicting harm. Adolescents mistakenly believe they are controlling the fate of a game character rather than destroying a real physical body, leading to a severe loss of a sense of reality.

5.3. Motivation and Reinforcement Stage

The continuation and escalation of behaviors depend on robust reinforcement mechanisms. Direct reinforcement, vicarious reinforcement, and self-reinforcement collaboratively weave the vicious cycle of "self-harm games". Vicarious reinforcement serves as the core driving force. Research by Sherman et al. (2016)confirms that "likes" on social media can substantially activate the reward circuits in adolescents' brains [5].Upon witnessing "high-level players" within the group display extreme imagery and garner immense attention and worship, adolescents are provoked with strong imitative motivations. In pursuit of equivalent status,

ISSN: 2688-8653

adolescents begin to replicate identical behaviors. Conversely, negative reinforcement establishes the maintenance mechanism. As the game progresses, individual motivation undergoes a fundamental reversal. The motive shifts from "seeking approval" to "avoiding punishment". To eliminate the fear of privacy exposure or ostracism, adolescents are forced to utilize self-harm to evade the cruel punishments of game manipulators. Fear-based compliance and vanity-based imitation intertwine. Ultimately, self-harm games degenerate from exploratory attempts into an inescapable, compulsive behavior.

6. Discussion and Implications

Based on the generative mechanisms delineated above, a systematic intervention strategy is proposed encompassing three dimensions-environment, cognition, and behavior-aimed at severing the interactive pathways of the vicious cycle of "self-harm games."

6.1. Environmental Regulation

At the level of environmental intervention, the core task lies in purifying the external milieu that induces observational learning. Regulatory authorities should collaborate with online platforms to establish a dynamic monitoring and circuit-breaker mechanism targeting invisible variant terms. Once a high frequency of specific "argot" triggers is detected, the system should immediately downgrade the recommendation weight of related content, thereby blocking the algorithmic distribution of harmful information at its source. Concurrently, platform algorithms must actively intervene to break the "echo chamber effect." For users who have previously browsed risk-related content, the system should mandatorily introduce counter-corrective information, such as psychological assistance resources or recovery narratives, utilizing shifts in environmental cues to distract and reallocate the negative attention directed at adolescents.

6.2. Cognitive Restructuring

At the level of cognitive restructuring, the focus of educational interventions should pivot from traditional preaching to a profound deconstruction of the game's manipulative logic, thereby enhancing adolescents' psychological fortitude. Schools should develop specialized media literacy curricula to dissect the psychological traps behind "task grading" and the legal invalidity of "death contracts," assisting adolescents in seeing through the facade of "cool challenges" to grasp their maliciously manipulative essence. Furthermore, educators must focus on cultivating adolescents' "digital resilience." When facing the threat of "participate or be kicked out" within online communities, this resilience enables them to recognize the threat merely as a negative reinforcement tactic based on fear, thereby equipping them with the psychological energy to resist cyber ostracism and proactively withdraw [4].

6.3. Behavioral Modification

At the level of behavioral modification, intervention strategies aim to construct positive vicarious reinforcement mechanisms and facilitate a return to real-world social interactions. Cyber regulators and educational departments should proactively occupy the online sphere by cultivating influential network leaders as positive role models. By launching theme-based online activities such as "Life Resilience Challenges," they can utilize positive feedback-such as "likes" and "follows"-to stimulate healthy imitative behaviors among adolescents. Simultaneously, schools must commit to building strong relational ties in the real world. By forming offline peer-support groups or organizing high-arousal cultural and sports activities, they can satisfy adolescents' psychological need for thrill-seeking while utilizing the sense of belonging in real life to dilute the allure of virtual groups, ultimately providing a safe, alternative outlet for self-harm tendencies.

ISSN: 2688-8653

7. Conclusion

This study aims to transcend traditional, single-factor psychopathological attributions by systematically analyzing adolescent "self-harm games" through the lens of Bandura's Triadic Reciprocal Determinism. The analysis reveals that this aberrant phenomenon is not merely an act of individual imitation, but rather the product of a sustained, vicious interaction among environmental factors (algorithmic echo chambers and anonymous communities), cognitive processing (symbolic translation and performative dissociation), and behavioral reinforcement (vicarious reinforcement and coercive cycles). Addressing these generative mechanisms, this paper proposes a multidimensional, collaborative intervention strategy encompassing environmental regulation, cognitive restructuring, and behavioral modification, thereby providing theoretical support and practical pathways for halting the proliferation of such extreme cyber subcultures.

Despite constructing a comprehensive theoretical integrated model, this study acknowledges certain limitations. Primarily relying on conceptual analysis and literature synthesis, the current research lacks validation from large-scale empirical data. Future research should employ quantitative surveys or longitudinal studies to empirically test the weights and causal relationships among the environmental, cognitive, and behavioral variables. Furthermore, cross-cultural comparative studies could elucidate the nuances of how "self-harm games" disseminate across varying socio-cultural contexts. Ultimately, combating adolescent cyber self-harm is a protracted, systematic endeavor that necessitates continuous societal vigilance and multisectoral coordinated intervention.

References

[1] F. Lv, Z.Y. Wang and Z.H. Song: Analysis of the characteristics of victims in internet-abetted suicide cases and preventive measures: Taking the "Blue Whale" death game as an example, Legal System and Society, (2020) No. 17, p.36-37. (In Chinese)

[2] H.D. Chen and L.D. Liu: Can collective anonymity evade cyberbullying? Taking the "Momo" group as an example, Journal of Shanghai Jiao Tong University (Philosophy and Social Sciences), Vol. 33 (2025) No. 1, p.78-91. (In Chinese)

[3] Y.F. Wang: The nihilism and reconstruction of the meaning of life among adolescents: Analyzing the Blue Whale game, Education Observation, (2019) No. 30, p.68-69. (In Chinese)

[4] Z.T. Zhu and S.S. Shen: Digital resilience education: Empowering students to grow happily in an increasingly complex world, Modern Distance Education Research, Vol. 32 (2020) No. 4, p. 3-10. (In Chinese)

[5] L.E. Sherman, A.A. Payton, L.M. Hernandez, et al.: The power of the like in adolescence: Effects of peer influence on neural and behavioral responses to social media, Psychological Science, Vol. 27 (2016) No. 7, p. 1027-1035.

[6] A. Khasawneh, K. Chalil Madathil, E. Dixon, et al.: Examining the self-harm and suicide contagion effects of the Blue Whale Challenge on YouTube and Twitter: Qualitative study, JMIR Mental Health, Vol. 7 (2020) No. 6, p. e15973.

[7] S. Mancone, G. Celia, A. Zanon, et al.: Psychosocial profiles and motivations for adolescent engagement in hazardous games: The role of boredom, peer influence, and self-harm tendencies, Frontiers in Psychiatry, Vol. 16 (2025), p. 1527168.

[8] A. Bandura: Social cognitive theory for personal and social change by enabling media, Entertainment-Education and Social Change: History, Research, and Practice, (2004), p. 75-96.