

The Logical Mechanism and Countermeasures of Online Group Polarization from the Perspective of Social Identity Theory

Zhu Yuqian¹, Zhang Shasha²

¹School of Journalism and Communication, Shanghai University, Shanghai, People's Republic of China

²(Corresponding author) Publishing College, University of Shanghai for Science and Technology, Shanghai, People's Republic of China

Email address: baohsan29@163.com

Abstract—With the rapid development of internet technology, cyberspace has become a crucial platform for social interaction. The emotional contagion characteristic of the post-truth era has also turned cyberspace into a breeding ground for online group polarization. From the perspective of Social Identity Theory, individual cognitive bias, group emotional contagion and social comparison are the key contributing factors to online group polarization, which is essentially a process of group interaction and individual identity construction. On this basis, this paper proposes countermeasures including strengthening media literacy education, optimizing mainstream algorithms, and constructing an online public sphere, so as to foster rational discussion and constructive communication in cyberspace.

Keywords— Social Identity Theory; Online Group Polarization; Logical Mechanism; Countermeasures.

I. INTRODUCTION

In the digital age, the Internet has become an important platform for social interaction and is reshaping people's social modes and cognitive structures. According to data from the 57th Statistical Report on the Development of China's Internet Network released by the China Internet Network Information Center (CNNIC), the number of netizens in China reached 1.125 billion as of December 2025, with the Internet penetration rate exceeding 80%.[1] With the rise of social media and the prevalence of algorithmic recommendation, individuals' behaviors and attitudes in cyberspace are increasingly influenced by groups, which gives rise to the phenomenon of online group polarization.

"Group polarization" is a concept put forward by scholar James Stoner to describe extreme phenomena of group behavior. It means that the inherent tendencies within a group are continuously amplified through mutual interaction, driving original opinions toward extremism: those who are originally conservative become more conservative, and those who are originally radical become more radical. [2] As an extension of group polarization, online group polarization and its negative impacts have become a key focus in the current governance of online public opinion. Sociologist Cass R. Sunstein said: "Like-minded groups communicate and deliberate with one another in the world of the internet and new communication technologies. In the end, their views remain essentially unchanged in essence, yet become more extreme in

form." [3] Gustave Le Bon referred to such irrational collectives as "the crowd" and established a theoretical system for understanding group events. [4] Through systematic reasoning and holistic cognition, he revealed the inherent nature that crowds are incapable of rational reasoning but eager to take action. [5] However, in the digital society, compared with Le Bon's exposition, online group polarization is characterized by conscious self-selection and unconscious collective behavior. This process involves not only interactions among groups but also individual factors such as social identity.

Social Identity Theory provides a powerful theoretical framework for explaining online group polarization. The theory holds that an individual's self-concept consists of not only personal identity but also the identity of the groups to which they belong. Individuals tend to distinguish themselves and their in-group from outsiders, on the basis of which in-group favoritism and out-group prejudice arise. In the online environment, such a sense of identity is further amplified. In the pursuit of identity and a sense of belonging, individuals tend to excessively reinforce the views of their own group while rejecting those of other groups, which further intensifies online group polarization. Adopting the framework of Social Identity Theory, this paper explores the logical mechanism of online group polarization and proposes a series of countermeasures. It intends to offer theoretical support and practical guidance for the sound development of cyberspace in the communication field, as well as to maintain social harmony and stability.

II. LITERATURE REVIEW

(1) Research Status of Online Group Polarization

Online group polarization has attracted sustained academic attention. Numerous scholars have investigated the formation mechanisms and influencing factors of group polarization mainly from three perspectives: communication studies, psychology and sociology. Wang Gensheng identified four manifestations of group polarization in online public opinion, namely unipolar aggregation, bipolar splitting, multipolar fragmentation and zero-polar dilution. He further analyzed factors affecting online group polarization including environment, duration, public attention and post popularity,

whose joint effects shape the trajectory of public opinion intensity.[6]

Against the backdrop of the "post-truth" era, Zhang Aijun and other scholars pointed out that public rationality is declining while emotional sentiment is intensifying. As a result, people fail to analyze issues with critical thinking, which impairs their capacity for independent and dialectical reasoning. [7]They proposed that governments should implement digital governance, strengthen media self-regulation, and improve the public's media literacy.

In his work *Infotopia: How Many Minds Produce Knowledge*, Cass Sunstein argued that although the Internet offers abundant information choices and convenient search tools, people tend to focus only on information and entertainment content of their own preference. In the long run, individuals will fall into the predicament of filter bubbles.[8]

Against the context of contemporary new media, the application of algorithms and big data technologies enables media platforms to accurately capture user preferences. Feng Gang et al. noted that with the extensive adoption of big data technologies by media in recent years, a large number of netizens have found themselves trapped in the so-called "filter bubble" phenomenon.[9]

(2) Research Status of Social Identity Theory

At the individual level, based on the well-known minimal group paradigm, Tajfel put forward the basic definition of social identity theory in 1978: "the individual's knowledge of his membership of a social group (or groups) together with the emotional significance attached to that membership." [10]At the individual level, social identity is reflected in people's perception of self-identity. It includes not only individuals' recognition of their membership in social groups, but also their acceptance and identification with the culture and values of the groups they belong to. As an inherent emotional need, social identity originates from the fundamental human desire to build connections with others and gain social recognition and acceptance.

Domestic scholars mainly focus on social identity issues among specific groups. For instance, Wang Zhonghui, Tong Huijie and Cheng Meng conducted questionnaire surveys among migrant children in Beijing and concluded that social identity exerts a significant influence on children's school adjustment. [11]Gao Wenjun, Yang Yiyin and other researchers carried out investigations among ethnic minority groups to explore structural types of dual social identity consisting of ethnic identity and national identity, and reveal the interactive relationships between multiple identity constructs.[12] Liu Jiawei implemented questionnaire research on online fan communities and found that strong in-group identity within communities tends to exacerbate intergroup antagonism and trigger extreme verbal conflicts within fandom circles.[13]

These studies demonstrate that social identity theory has been widely adopted by domestic researchers to analyze various groups and explore its functions in education and social adaptation. At present, few scholars employ social identity theory as an analytical framework to examine group

polarization. Therefore, studying group polarization from individual and group perspectives and proposing corresponding countermeasures helps broaden relevant research horizons and enrich the theoretical system of social identity theory.

III. ANALYSIS OF THE LOGICAL MECHANISM OF ONLINE GROUP POLARIZATION

(1) Individual Level: Cognitive Bias and Identity Construction

Cognitive bias refers to individuals' irrational judgments and decisions arising from psychological, emotional and cognitive factors during information processing.[14]In the online environment, due to the massive volume and complexity of information, people are more susceptible to preconceived notions and preferences. They selectively access, interpret and process information, thereby forming and reinforcing their existing beliefs and attitudes. This phenomenon is particularly prominent on social media and has been further intensified in the internet era. Users are often confined within information bubbles of their own interests, lacking exposure to and engagement with heterogeneous information, which ultimately exacerbates the phenomenon of online group polarization.

The construction of individual identity is a core concept in Social Identity Theory. It concerns how individuals define the self, sense of belonging, and their relationships with social groups.[15]In cyberspace, individuals continuously construct and reconstruct their social identity through interaction and communication with others. For example, self-presentation and interactive feedback on social media constitute a direct process of individual identity construction. Taking Weibo as an example, users display their lives and viewpoints by posting polished photos, texts and videos, thereby shaping an idealized self-image. Meanwhile, through feedback such as likes, comments and shares from their social circles, users perceive and confirm their status and sense of identity within the social network.

There exists a complex interactive relationship between cognitive bias and identity construction. On the one hand, cognitive bias influences how individuals interpret information related to group identity, making them more inclined to accept views consistent with their own while ignoring information that contradicts their self-identification. On the other hand, the process of individual identity construction also shapes the direction and intensity of cognitive bias. To maintain their social identity and better integrate into a certain group, individuals tend to exhibit highly relevant cognitive biases. This two-way interactive relationship reinforces online group polarization and drives the opinions of individuals and groups toward greater extremism and confrontation.

(2) Group Level: Group Emotion and Social Comparison

In the process of group interaction, emotional contagion means that individuals' emotions and behaviors are influenced by those of other group members, leading them to generate similar emotional states and behavioral tendencies. In the

context of online group polarization, emotional contagion is particularly prevalent. Driven by the mechanism of emotional contagion, extreme emotional expressions on social media platforms spread at an astonishing speed, thereby pushing group emotions toward radicalization. For example, as the Tangshan barbecue restaurant assault incident spread widely across social media, public sentiment was quickly aroused through emotional contagion. Emotions such as anger, fear and a sense of insecurity spread rapidly among internet users and generated strong emotional resonance. On social media, various groups expressed their views and sentiments on the incident by posting articles, leaving comments and sharing content, which further intensified emotional expression. With the spread of emotional contagion, group emotions became radicalized, giving rise to online group polarization. On the one hand, public anger and condemnation toward the perpetrators reached an unprecedented level, and calls for severe legal punishment grew increasingly loud. On the other hand, influenced by emotional extremism, some people escalated the incident into gender confrontation, aggravating opposition and division among different groups and eventually forming obvious group polarization. In addition, the case of the pink-haired girl who committed suicide after being subjected to online violence fully demonstrates the profound impact of emotional contagion on group polarization.

Social comparison theory holds that individuals evaluate their social status and self-worth by comparing themselves with others. In the online environment, group members tend to compare themselves with those who hold similar viewpoints. Such a comparison process may intensify consensus within the group, trigger rejection and hostility toward the out-group, and consequently strengthen or weaken group cohesion.

Emotional contagion and social comparison interact with each other in online group polarization and jointly drive its development. Emotional contagion reinforces the outcomes of social comparison, causing group members to generate stronger emotional reactions once they perceive differences from out-groups. It may also intensify intra-group emotional contagion. When individuals feel that the status of their in-group is under threat, they become more susceptible to emotional influences within the group, thereby enhancing group cohesion and the degree of exclusion toward out-groups.

(3) Social Level: Algorithmic Recommendation and Information Cocoons

The embedded algorithmic mechanisms and information cocoons of online platforms constitute a major factor leading to online group polarization. The information cocoon effect describes the phenomenon of information isolation, in which users are only exposed to content consistent with their personal stances in cyberspace. For instance, on personalized news recommendation platforms such as Toutiao, users are often trapped in information bubbles. By analyzing users' behavioral patterns and preference trajectories, algorithmic recommendation mechanisms customize personalized content and exacerbate the formation of information cocoons to a certain extent. Reinforced by technical calculation, such

recommendation mechanisms consolidate users' inherent viewpoints, strengthen the echo chamber effect, and confine users within a narrow scope of personalized information.

Algorithm-driven information screening mechanisms make it easier for users to interact with group members who hold similar views. Based on the principles of selective exposure and the mechanism of emotional contagion, individuals further form stronger shared emotions and opinions within the group, transforming the individual information cocoon into a group information cocoon. There exists a dynamic interactive relationship between information cocoons and algorithmic recommendation. Information cocoons feed user preference data and behavioral trajectories back to algorithmic recommendation systems. In turn, fueled by such data, algorithmic recommendations further deepen the isolation effect of information cocoons. This two-way interaction jointly exacerbates the phenomenon of online group polarization. For example, during the 2016 U.S. presidential election, algorithmic recommendation systems were hijacked for political purposes to spread fake news and misleading information, building barriers between misleading content and factual information and thus forming information cocoons. The data generated within information cocoons further expanded the coverage and intensity of algorithmic recommendations. Under the combined influence of information cocoons and algorithmic mechanisms, social and political opinions became increasingly radicalized. As a result, users are hardly exposed to authentic information and diverse perspectives, falling into a state of pluralistic ignorance accompanied by group polarization.

IV. COUNTERMEASURES AGAINST ONLINE GROUP POLARIZATION

(1) Strengthen Online Literacy Education

From the perspective of Social Identity Theory, improving online literacy education means enhancing the public's ability to acquire, understand and utilize information in the online environment, which is of great significance for resisting cognitive bias at the individual level. Strengthening online literacy education can improve the public's ability to identify information, curb the spread of extreme viewpoints and misinformation, and thereby reduce the occurrence of group polarization.

In the information-exploded online environment, individuals need to possess the ability to screen and evaluate the authenticity of information so as to avoid being misled by false or biased content. Education authorities and media organizations can cooperate to carry out relevant training and publicity activities and raise netizens' awareness of suspicious information. It is essential to cultivate critical thinking and diverse perspectives. The development of critical thinking enables individuals to examine issues from multiple dimensions, thereby preventing cognitive bias and ideological extremism caused by a single-sided viewpoint.

The practical path of online literacy education should be built on an integrated system covering policy formulation, platform governance and school education. As an important venue for talent cultivation, schools should incorporate online

media literacy into formal curricula or establish public online platforms to provide flexible learning resources, enabling the public to acquire relevant knowledge of online literacy anytime and anywhere. Furthermore, online literacy education should not be confined merely to technical skills; it should also include emotional education and value cultivation. Educators ought to guide the public to understand and respect diverse cultures and develop strong empathy, so as to reduce the likelihood of online group polarization.

(2) Optimize the Mainstream Algorithmic Recommendation Mechanism

Cyberspace is the shared spiritual home of humanity. Mainstream culture constitutes the foundation of social ideology and serves as an effective bulwark against online group polarization. In the digital age, people are immersed in all kinds of entertainment offered by modern technology and have long lost the capacity to perceive self-contradiction.[16]Addressing the problem of online group polarization is a crucial step in governing cyberspace.

Secondly, optimizing mainstream algorithmic mechanisms requires enhancing algorithm transparency and explainability, while standardizing the accountability and ethics of algorithmic recommendation. The algorithmic black box creates contradictions between the public and technology. Improving algorithm explainability can strengthen public trust in technology and enable audiences to view diverse viewpoints more rationally. Meanwhile, algorithmic recommendation systems must abide by principles of accountability and ethics. Algorithms should not only pursue commercial benefits but also take social value and social welfare into account. Algorithm design should avoid exacerbating social division and polarization; instead, it ought to promote information diversity and balanced perspectives. Platforms should also establish ethical review and punishment mechanisms to prevent algorithms from recommending content that deviates from mainstream values and ethical norms. Finally, the optimization of technical approaches not only needs to incorporate mainstream ideology as a spiritual foundation, but also introduce diversity indicators. By learning from users' behavioral feedback, platforms can dynamically adjust recommendation strategies to better serve users' long-term interests, align with mainstream social values, and contribute to overall social well-being.

(3) Constructing the Online Public Sphere

In the digital internet era, the communication patterns of traditional society have undergone profound changes. Based on personal interests and values, people have formed new social networks online — online cliques. Free from geographical constraints, these online cliques enable continuous and multi-dimensional interaction through the time-space transcending communication mode provided by the internet. Nevertheless, such networks tend to evolve into closed circles. Within this closed environment, group identity often overrides individual identity, and is even mistakenly regarded as individual identity in some cases. This phenomenon may cause individuals to abandon independent

thinking under group pressure, thereby inhibiting the development of critical thinking.

Habermas's concept of the public sphere was put forward in the 1960s, referring to an ideal space where citizens can freely gather, discuss, and debate public affairs. Meanwhile, it must be acknowledged that the public sphere depicted by Habermas is an idealized model. In reality, the public sphere is often influenced by political, economic, social and other factors. At times, the so-called public sphere may even become a breeding ground for online group polarization. For instance, Weibo has not evolved into an ideal venue for rational deliberation; instead, it has turned into a mixed platform open to unrestricted speech by all users, which frequently gives rise to online group polarization. Rooted in the classical notion of the public sphere, the online public sphere emphasizes a communication model adapted to the digital age within the internet environment. It provides a platform for individuals with diverse backgrounds and viewpoints to engage in open and equal dialogue, and thus plays a vital role in responding to and mitigating online group polarization.

To foster rational discussion and curb the interaction between emotional contagion and social comparison mechanisms, the online public sphere needs to design effective mechanisms and strategies to encourage citizens to participate in public deliberation on social affairs. This includes ensuring the inclusiveness of discussions, implementing the aforementioned new algorithmic mechanisms, and establishing standardized rules and institutions. Meanwhile, the online public sphere should always uphold the principle of platform neutrality, strictly uphold the bottom line of mainstream core values, and set up a user feedback mechanism to adjust relevant algorithmic mechanisms in real time. Finally, targeted early warning mechanisms should be established on platforms to address severe online group polarization. Through rational and timely intervention, platforms can mitigate radical conflicts and contradictions between opposing groups.

V. CONCLUSION

From the perspective of social identity theory, this paper explores the logical mechanism of online group polarization and corresponding countermeasures. As a product of the changes in social interaction patterns in the digital era, online group polarization is closely related to individual cognitive bias, group emotional contagion, and social comparison. Its formation is not only the outcome of group interaction but also a process of individual identity construction. At the individual level, cognitive bias inclines internet users to selectively access and process information, thereby reinforcing their existing beliefs and attitudes. At the group level, emotional contagion and social comparison function in group interactions, driving group members' opinions and behaviors toward extremism. At the social level, algorithmic recommendation and the information cocoon effect further aggravate online group polarization. Users are often confined within information bubbles and rarely exposed to diverse perspectives. To cope with online group polarization, this

paper proposes three countermeasures: strengthening online literacy education, optimizing mainstream algorithmic mechanisms, and constructing the online public sphere. This research not only provides theoretical support for understanding online group polarization, but also offers strategic references for practical governance.

REFERENCES

- [1] CNNIC. (n.d.). The 57th Statistical Report on China's Internet Development. Ebrun E-Commerce Research Center.
- [2] Xu, S. F. (2012). Network public opinion: A new symptom of urban public security crises. *Nanjing Journal of Social Sciences*, 4, 64-69.
- [3] Sunstein, C. R. (2003). *Republic.com 2.0* (Huang, W. M., Trans.). Shanghai People's Publishing House. (Original work published 2002) (p. 47)
- [4] Yuan, G. F., & Liu, C. P. (2022). Defining "the masses": A sociological examination of the dissemination of The Crowd in China. *Journalism & Communication Research*, 1, 56-78.
- [5] Le Bon, G. (2017). *The crowd: A study of the popular mind* (Feng, K. L., Trans.). Central Compilation & Translation Press. (Original work published 1895) (p. 69)
- [6] Wang, G. S. (2012). Dynamic model and simulation analysis of group polarization of network public opinion. *Journal of Intelligence*, 31(3), 20-24.
- [7] Zhang, A. J., & Liang, S. (2019). Negative effects and avoidance strategies of online group polarization. *Academic Forum*, 4, 75-83.
- [8] Sunstein, C. R. (2008). *Infotopia: How many minds produce knowledge* (Bi, J. Y., Trans.). Law Press China. (Original work published 2006)
- [9] Feng, G., & Huang, Y. L. (2021). Causes and characteristics of group polarization of college students' online public opinion. *Studies in Ideological Education*, 9, 93-98.
- [10] Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. Academic Press.
- [11] Wang, Z. H., Tong, H. J., & Cheng, M. (2016). The influence of social identity on school adaptation among migrant children. *Chinese Journal of Special Education*, 3, 52-57.
- [12] Gao, W. J., Zhao, Z. Y., Yang, Y. Y., et al. (2013). The structure and influence of dual social identity: National and ethnic identity—Evidence from a survey of Han and ethnic minority residents in Yunnan. *Journal of Yunnan Normal University (Humanities and Social Sciences Edition)*, 45(5), 64-75.
- [13] Liu, J. W. (2024). Collective behaviors of online fan communities from the perspective of social identity theory. *Journal of Social Science of Hunan Normal University*, 53(3), 130-140. <https://doi.org/10.19503/j.cnki.1000-2529.2024.03.014>
- [14] Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- [15] Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Brooks/Cole.
- [16] Postman, N. (2015). *Amusing ourselves to death* (Zhang, Y., Trans.). CITIC Press Group. (Original work published 1985) (p. 32)