Proceedings of the 2nd International Conference on Global Politics and Socio-Humanities
DOI: 10.54254/2753-7048/75/2024.17839

Analysis on the Problems of Internet Conflicts and Countermeasures

Yongkang Su[1,a,*]

[1]*Changwai Bilingual School, Changzhou, 213000, China*
a. sszsdr@163.com
**corresponding author*

Abstract: Polarization and conflict online have grown as one of the most pervasive issues that characterize the contours of the digital environment, with serious risks to social cohesion, democratic discourse, and well-being. This paper thus seeks to trace those very root causes of online conflicts and the dynamics of the escalation process into group-based polarization. The paper considers a proposed model of how personal disagreements between two individuals scale up to large groups of conflicts with the help of cognitive biases, anonymity, and algorithmic reinforcement on digital platforms. Possible policy interventions at each stage of the escalation of conflict are discussed: from digital monitoring and public education about cognitive biases to online mediation programs and changes in algorithms to facilitate cross-cutting interactions. These findings suggest that only multilevel interventions that combine technological, social, and psychological approaches to mitigate conflict while preserving freedom of expression are effective strategies. The paper thus concludes that long-term solutions include changes in culture in building empathetic individuals and critical thinkers, not just immediate moderation. This gives a basis to governments and policymakers on how to understand the complexity and nuances involved in online conflict, thus forming the basis on which more balanced and inclusive digital environments could be developed that foster healthy dialogue and reduce polarization.

Keywords: Online polarization, Group conflict, Cognitive biases, Algorithmic interventions, Digital moderation.

1. Introduction

Online polarization and conflicts are increasingly prevalent, which over the years has caught the escalating concern [1,2]. The growing adversarial dynamics and entrenchment of extreme views on digital platforms go to pose a big menace to individuals, communities, and societies at large. The widespread occurrence of online conflicts and polarization can be highly risky to individual and collective health. Of particular concern is the potentiality of escalating self-injury and suicidal practices among vulnerable youth groups. Research studies have noted that the use of social media networks may increase the risk of self-injury and attempted suicide [3]. The viewing of online self-injury content tends to normalize these behaviors and subsequently influence imitation practices [4]. Moreover, such dissemination can involve highly detrimental consequences: hate speech is not only designed to attack specific individuals or groups but also engenders an overall atmosphere of intolerance and discrimination upon which it depends for its foundation [5]. This can weaken social

DOI: 10.54254/2753-7048/75/2024.17839

cohesion and damage the base on which any pluralistic, democratic society exists. Another risk is that governments will attempt to place restrictions on specific "harmful" types of speech, which may violate basic freedoms of expression [6]. As a result, this paper aims to build an entirely new model to show the origin and cause of online conflicts and polarization. And according to this model, some methods and solutions could be listed to deal with those problems.

2. Literature Review

Polarization and conflict online are common realities today, the result of a multi-directional nexus of technological, psychological, and social factors. The first root cause is the fact that political discourse is hugely predisposed to polarize and radicalize on digital platforms. Social media algorithms often reinforce users' prevailing attitudes by recommending content reflective of their political orientation and creating "echo chambers" in which extreme opinions get magnified [7,8]. It reduces the likelihood that people are exposed to different viewpoints and reduces the potential for constructive conversation across ideological divides.

Secondly, social identity and in-group versus out-group perceptions are more salient online, thereby maintaining and perpetuating affective polarization by building up more negative feelings toward others holding views different from their own. The anonymizing and direct interpersonal interaction-reducing nature of digital platforms can reduce empathy and impede constructive handling of disagreement. People might be less inhibited in expressing hostility when interacting with others under a perceived cloak of anonymity and thus engage in "flaming" or other forms of trolling [9,10]. The unabated spread of misinformation and "fake news" through social media, coupled with a problem that only cements polarized beliefs and erosion in shared reality, just doesn't seem to stay that course. Often, misinformation toys with people's biases and emotions already in their heads, which again makes them all the more believable and sharable. This easily fosters parallel realities when people hold fundamentally different views on reality and when productive dialogue and compromise are increasingly impossible to realize [11,12].

The antagonistic dynamics indeed have more often than not ballooned on the digital platforms due to an absence or inefficiency in strategies to moderate and manage the conflict. Moreover, design choices on the platforms-in particular, the algorithms that maximize engagements, and nearly friction-free content sharing-can incentivize inflammatory and divisive content. Moreover, moderators hardly manage to deal with the amount of abusive content, and their reactions may sometimes be viewed as one-sided or too heavy-handed, thus making resentment even bigger [13,14]. There is a need to address the multilevel interventions toward technological, psychological, and social drivers to mitigate online polarization and conflict. That would involve the redesigning of recommendation algorithms that foster exposure to diverse perspectives, nurturing digital media literacy for enabling critical assessment of information found online, allowing cross-cutting interactions that enable one to empathize with others and understand their perspective, and strong mechanisms of conflict resolution that help in damping tensions [15-17].

Instead, one of the more promising approaches involves using "digital nudges" to facilitate more constructive online behavior. Examples include requiring users to consider others' perspective before posting a comment or embedding friction into the sharing of potentially inflammatory content. Indeed, such interventions may enlist insights from behavioral science in order to nudge people toward more prosocial and less polarized interactions [18,19].

More importantly, these will be achieved by deepening the levels of digital media literacy and critical thinking to better position people to navigate the online information space. Helping people spot misinformation, how algorithms work, and how to hold productive discussions would enable us to dampen some of the drivers of online polarization and conflict [11,20]. Second, cross-cutting interactions could help engender empathy across ideological divides. Cross-cutting activities bring

DOI: 10.54254/2753-7048/75/2024.17839

diverse perspectives together, either through online structured discussions or through in-person deliberative forums, and therefore can play a part in reducing affective polarization through mutual understanding. Such interventions humanize "the other," create space for real dialogue to flourish, and thus may counter the rather frequent appearance of tribalism and hostility seen on the internet [15,17].

Overall, we could see that there have been a lot of methods and solutions that focus on the topic of how to deal with online conflict. However, there are not many essays focus on the how to deal with it from the society and government aspects. Thus, a model to analyse the origin of online conflicts is necessary for this paper to investigate in order to give a reference for government and society to determine and promote new policies to improve the bad situation of online conflicts.

3. Model and Analysis

This paper creates a model of the creation of online conflicts. There are totally four processes for this model. Each arrow explains a shift of the aggressive emotion. It starts with a conflict between two individuals A and B and finally leads to a conflict between Group A and Group B which is the final result, online conflict. This model only could be used for online conflict because the Internet has a key feature of "public".

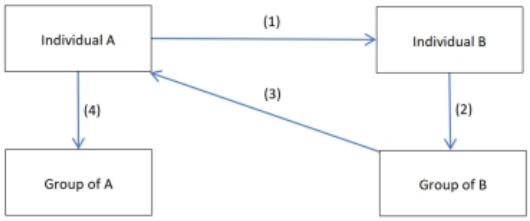

Figure 1: The model of creation of group polarization.

Figure 1 shows the cause of online group conflict. For process (1), it may be some personal reasons for a conflict between two people individual A and individual B. For instance, somebody does not like other's behavior. As a result, individual A comments some aggressive words toward individual B. However, usually, the aggressive words will not only be specific to the individual itself.

Process (2) shows the effect of the shift of aggressive emotion. It means that the aggressive emotion from individual A towards individual B could commonly shift to the group that individual B stays in. There is also the "ultimate attribution error," another type of cognitive bias wherein people generalize an individual's behavior to the larger group to which they belong [21,22]. It occurs when an individual exhibits an act that people attribute to the inherent characteristics of their group, rather than to situational factors that contributed to such behavior [23]. The result is that an unwanted action committed by one member of an ethnic or religious group is quickly generalized to the whole group as representative, without determining the circumstances that influenced the particular behavior of the individual concerned. This type of over-generalization serves only to create malignant stereotypes and further polarize prejudices [24], as it disregards individual differences within groups. Indeed, studies within social psychology have elucidated that people are likely to attribute good behaviors of in-group members to internal, stable factors while attributing negative behaviors to external, unstable ones. Meanwhile, for out-group members, positive behaviors tend to be overlooked, while negative behaviors are attributed to internal, stable factors, hence perpetuating negative perceptions. This one-sidedness distorts the presentation of individual actions and, moreover, is socially defective as it disadvantages and discriminates against minority groups, thereby further injuring their chances of being seen and valued as individuals in their own right. Dovidio and his groups [25] argue that this form of bias has significant practical effects: it reifies negative stereotypes and widens social divides.

DOI: 10.54254/2753-7048/75/2024.17839

As a result, the group will have a negative impression of individual A as the arrow (3) shows and thus this negative aggressive emotion will upgrade into the aggressive emotion towards Group A from Group B as process (4) shows. It is the same reason for the shift of emotion of process (2). Finally, the conflict between two groups, Group A and Group B, exists.

4. Proposed Policies

The government could only prevent online conflict if interventions in each stage of the escalation process are made as identified within the model. In other words, each policy recommendation should be analyzed based on the potential it holds to reduce conflict, avoid escalation, and achieve a more constructive atmosphere online.

4.1. Process 1: The Aggressive Behavior of Individual A Against Individual B

Policy Recommendation: Digital Conduct Regulation and More Efficient Online Monitoring

It is, therefore, upon the government to set very stringent regulations that compel social media to install a real-time moderation system able to detect injurious language successfully. Such systems, driven by AI, would automatically identify aggressive behavior-for example, insults or threats-and immediately notify moderators to take the proper measures: warnings, temporary suspension, or, in case of recidivism, conflict resolution courses. Besides this intervention, governments could push for transparency policies to make sure regular reports are published on the effectiveness of such interventions.

The success of this policy depends on the ability to balance moderation and freedom of speech. While this might prevent personal attacks at the individual level, unbridled moderation might repress genuine discussion. To this end, it is important that AI systems be developed to tell the difference between harmful aggression and heated, yet permissible, debate. This will be ensured through routine audits by independent bodies. This policy might greatly reduce the ignition of conflicts, but it would only work based on the public's perception that this was being carried out in a nondiscriminatory manner. It requires trust in the equity of its enforcement. According to [12].

4.2. Process 2: Diffusion of Aggression from Individual B to Group B

Policy Recommendation: inclusive public education campaigns on cognitive bias and digital literacy. Second, this would require the governments to pursue extensive education campaigns informing people about types of cognitive biases involved in generalizing individual actions to whole groups-such as what has been termed the "ultimate attribution error" [22]. These would form part of the school curriculum and public media and government websites, thereby letting the general public know that biases feed online hatred based on groups. In addition to that, it is the role of the government to support programs on digital literacy that will allow users to learn how to critically evaluate the content online and challenge the stereotypes when necessary.

These campaigns are relevant to address profound psychological mechanisms which magnify individual conflicts into group-based tensions. Over time, acquainting people with the psychological underpinning of online behavior may reduce stereotyping and generalizing of negative behaviors. However, these campaigns can have a real impact only after a number of years; it consists of a cultural and social change. Success of this policy can be gauged through the results of public opinion surveys, along with monitoring the reductions in group-based online hostility [11].

4.3. Process 3: Group B's Negative Reaction against Individual A

Policy Recommendation: Government-Run Online Mediation Programs.

DOI: 10.54254/2753-7048/75/2024.17839

It means that federal, state, and local governments can establish a national online mediation system where certified mediators would intervene online in disputes at an early stage of escalation. Third-party mediators would provide a neutral venue where aggrieved individuals and groups can express their grievances and sort out their differences before it escalates into broader conflict. The government can develop a publicly financed platform through which users can report conflicts or aggression that can be forwarded for resolution to such mediators.

Online mediation would offer a new channel toward dealing with disputes, which overcomes some of the weaknesses of platform-led moderation that lacks nuance. Indeed, such a policy would be able to de-escalate incidents with a human approach and, in fact, far sooner. Scalability is tough with such programs, given that online conflicts happen fast and often. This would mean a great government investment in mediator training and infrastructure to handle massive volumes of disputes. Successful results could be measured by reduced recurrence of reported disputes by the same disputants [17].

4.4. Process 4: Escalation into Conflict Between Group A and Group B

Policy Recommendation: Facilitating Cross-Cutting Interactions and Algorithmic Interventions

Because of this, governments should compel social media platforms to change their algorithms in ways that will promote cross-cutting interactions, hoping these may prevent conflicts from crystallizing into entrenched group-based divides. Algorithms might be rewritten to favor diverse perspectives; users would show content drawn from outside the narrow ideological or social "echo chambers." Additionally, governments could incentivize platforms hosting online rooms where users from opposing groups can engage in structured dialogue with a moderator. Such forums would cultivate empathy and understanding across divides.

Algorithmic changes could, therefore, make a big difference in depolarizing and decreasing conflict between groups by breaking up insular environments that reinforce the most extreme views. Forcing exposure to diverse perspectives can undermine entrenched bias, creating greater empathy across opposing groups. However, if such exposure is coerced, any number of backlashes may occur, since entrenchment is likely unless the process is carefully handled. This policy's effectiveness should therefore be monitored through changes in polarization metrics, such as the diversity in use of content by any given user and the frequency of hostile interaction between groups [1]. In addition, user feedback about dialogue forums provides an indication of how well meaningful conversations are being fostered.

5. Conclusion

Together, these three threaten to unleash a ferment of social cohesion, democratic discourse, and well-being. The model developed in this paper captures how personal conflicts scale into group-based online confrontations, while personal biases, anonymity, and extreme views are amplified by algorithms. Increased digital monitoring, public education about cognitive biases, sponsored mediation by the government, and encouragement of cross-cutting interactions are the policy interventions that can be done in order to dampen the escalation of online conflict without compromising freedoms of speech and expression. However, there are a number of limitations in the present model and approach that need consideration in future research.

For example, it only partly explains how cultural, economic, and geopolitical factors come into play in the shaping of online behaviour across dissimilar contexts. Future research should tease out precisely how such structural forces cut across online dynamics in society. This is a problem because a majority of the literature reviewed within this paper has been conducted on Western digital platforms. This leaves rooms for expansion in order to understand online conflict even within non-Western societies, where social media uses and regulatory frameworks are considerably different.

DOI: 10.54254/2753-7048/75/2024.17839

This could be further improved methodologically by the inclusion of empirical data, for instance, case studies or user-behavior analysis, to substantiate the model proposed. Because no empirical testing has occurred, it is yet to be known whether these interventions are going to work. Therefore, quantitative and qualitative studies shall be undertaken in the future research to help tease out how the intervention works in practice and might be scaled effectively. Future research would therefore also need to consider the ethical dimensions of algorithmic interventions and digital nudging. That is, the sensitive balance between promoting healthy online discourse without undermining user autonomy.

As technology continues to evolve, the insight into the emergent platforms and new ways of communication-for example, VR and decentralized social networks-will be influencing future policy. This model completes the gaps and hence perfection from the original and enables both policymakers and researchers to better foster a healthy and inclusive digital environment.

References

[1] *Bail, C. A., Argyle, L. P., Brown, T. W., et al. (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 115(37), 9216-9221.*

[2] *Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater internet use is not associated with faster growth in political polarization among US demographic groups. Proceedings of the National Academy of Sciences, 114(40), 10612-10617.*

[3] *Mok, K., Jorm, A. F., & Pirkis, J. (2016). Suicide-related Internet use: A review. Australian & New Zealand Journal of Psychiatry, 50(8), 704-724.*

[4] *Arendt, F., Scherr, S., & Romer, D. (2019). Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults. New Media & Society, 21(11-12), 2422-2442.*

[5] *Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). Countering online hate speech. Unesco Publishing.*

[6] *Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. Harvard Law Review, 131, 1598.*

[7] *Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130-1132.*

[8] *Geschke, D., Lorenz, J., & Holtz, P. (2019). The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. British Journal of Social Psychology, 58(1), 129-149.*

[9] *Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017, February). Anyone can become a troll: Causes of trolling behavior in online discussions. In Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing, pp. 1217-1230.*

[10] *Rösner, L., & Krämer, N. C. (2016). Verbal venting in the social web: Effects of anonymity and group norms on aggressive language use in online comments. Social Media+ Society, 2(3), 2056305116664220.*

[11] *Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science advances, 5(1), eaau4586.*

[12] *Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, 116(7), 2521-2526.*

[13] *Matias, J. N. (2019). Preventing harassment and increasing group participation through social norms in online communities. In Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-23.*

[14] *Seering, J., Kraut, R., & Dabbish, L. (2017). Shaping pro and anti-social behavior on Twitch through moderation and example-setting. In Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing, pp. 111-125.*

[15] *Bail, C. A. (2021). Breaking the social media prism: How to make our platforms less polarizing. Princeton University Press.*

[16] *Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. Health communication, 33(9), 1131-1140.*

[17] *Margolin, D. B., & Hancock, J. T. (2020). Advancing digital civility: An interdisciplinary framework for reducing online incivility. American Psychologist, 75(5), 702.*

[18] *Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. Political Behavior, 39(3), 629-649.*

DOI: 10.54254/2753-7048/75/2024.17839

[19] *Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychological science, 31(7), 770-780.*

[20] *Vraga, E. K., & Tully, M. (2019). News literacy, social media behaviors, and skepticism toward information on social media. Information, Communication & Society, 22(2), 203-219.*

[21] *Allport, G. W. (1954). The nature of prejudice. Addison-Wesley.*

[22] *Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. The Social Psychology of Intergroup Relations, 33(47), 74.*

[23] *Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. Personality and Social Psychology Bulletin, 5(4), 461-476.*

[24] *Hewstone, M. (1990). The ultimate attribution error? A review of the literature on inter-group causal attribution. European Journal of Social Psychology, 20(4), 311-335.*

[25] *Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. Journal of Personality and Social Psychology, 82(1), 62-68.*