Simen Owen Academic Proceedings Series

3rd International Conference on Education, Environment, Arts and Social Science (EEAS 2026)

Article

From Public Discourse to Affective Tribes: Mechanisms of Affective Communication and Moral Judgment in Social Media Gender Debates

Chenjun Zhao [1,*]

[1] College of Automation (College of Artificial Intelligence), Beijing Information Science & Technology University, Beijing, China

* Correspondence: Chenjun Zhao, College of Automation (College of Artificial Intelligence), Beijing Information Science & Technology University, Beijing, China

Abstract: In recent years, social media has fundamentally transformed the landscape of public discourse on gender issues. Instead of facilitating rational deliberation within a traditional public sphere, digital platforms have devolved into polarized battlegrounds characterized by affective tribalism. This study investigates the mechanisms driving the shift from rational public opinion to emotional side-taking and aggressive moral policing in online gender debates. Utilizing the highly polarized "Yang Li and JD.com" controversy as a primary case study, this research employs a mixed-methods approach, integrating computational sentiment analysis (NLP) and qualitative discourse analysis, to examine 42,000 highly engaged Weibo comments. The empirical findings reveal three key phenomena. First, sentiment analysis demonstrates a severe U-shaped (bimodal) distribution, providing quantitative evidence for the hollowing-out of the objective middle ground and the dominance of extreme emotions. Second, lexical mapping illustrates that users increasingly bypass logical argumentation, deploying stigmatizing gender labels as heuristic tools for tribal boundary-drawing. Third, statistical correlation reveals that content with high emotional volatility and absolute moral framing receives significantly higher user engagement. Theoretically, this paper argues that affective communication functions as a mechanism for identity politics, where expressing collective outrage confirms tribal loyalty and escalates minor frictions into systemic moral judgments. Furthermore, the study highlights the structural complicity of platform algorithms, which systematically reward polarized outrage to maximize the attention economy. To reconstruct a deliberative digital space, this paper calls for algorithmic accountability and the cultivation of affective digital literacy.

Keywords: affective publics; emotional tribalism; moral policing; sentiment analysis; gender discourse

Received: 12 April 2026
Revised: 19 May 2026
Accepted: 30 May 2026
Published: 04 June 2026

1. Introduction

In recent years, social media has fundamentally transformed the landscape of public discourse, particularly concerning gender issues. Initially envisioned as a digital Habermasian public sphere facilitating rational-critical debate, these platforms have increasingly devolved into highly polarized battlegrounds. Gender debates, inherently interwoven with complex social structures and personal identities, are now frequently characterized not by a mutual search for consensus, but by ideological fragmentation and deep-seated hostility [1]. This phenomenon marks a significant departure from traditional public opinion formation, shifting the paradigm toward what can be described as affective tribalism.

This profound shift raises urgent questions regarding the underlying mechanisms of contemporary digital communication [1]. The displacement of rational dialogue by emotional polarization indicates that affect, rather than logic, has become the dominant currency of online interaction. In the context of gender controversies, users increasingly bypass objective deliberation in favor of expressing collective outrage, seeking solidarity within echo chambers, and deploying aggressive moral judgments against perceived opponents. Consequently, nuanced social issues are often reductionistically reframed as binary moral conflicts. It is imperative to ask: How does affective communication catalyze this transformation from rational discourse to emotional side-taking? Furthermore, what specific mechanisms drive the escalation from basic emotional expression to rigid moral policing within these digital tribes?

To address these questions, this paper investigates the intricate interplay between affective communication and moral judgment in social media gender debates. By examining recent prominent gender-related controversies, this research employs a mixed-methods approach, integrating computational sentiment analysis with qualitative discourse analysis [2]. The primary objective is to map the trajectory of emotional contagion and uncover how specific affective states crystallize into rigid moral boundaries that divide online communities.

The significance of this study is twofold [3]. Theoretically, it advances the framework of "affective publics" by offering an empirical examination of how emotional tribalism operates at the intersection of human psychology and platform affordances. Practically, by demystifying the structural and emotional mechanics that fuel online gender antagonism, this research provides valuable insights for algorithmic governance, digital literacy, and the potential reconstruction of a healthier, more deliberative digital public space.

2. Theoretical Foundation and Literature Review

2.1. Paradigm Shift: From the "Public Sphere" to "Affective Publics"

Traditional communication studies have long relied on the normative concept of the "public sphere," which posits a space where equal participants engage in rational-critical debate to reach consensus. However, the utopian assumption of communicative rationality is increasingly obsolete in the contemporary digital ecology [4]. On platforms like Weibo or X (formerly Twitter), structural constraints such as character limits and algorithmic curation actively discourage sustained logical argumentation. In reality, modern digital gender debates are rarely driven by the "force of the better argument."

To address this theoretical gap, the concept of "Affective Publics" provides a more accurate paradigm. This framework argues that networked publics are mobilized and connected primarily through expressions of sentiment rather than rational consensus. In the context of digital gender controversies, truth and objective facts often become secondary to the "affective resonance" an event triggers. For instance, when a gender-related conflict surfaces, the immediate public reaction is rarely a call for evidence; instead, it is an instantaneous outpouring of collective empathy or indignation. The network is thus bound together by shared feelings, transforming the public arena into a theater of affective performance [5].

2.2. "Affective Tribes" and Networked Group Identity

This reliance on shared emotion fundamentally alters how online group identities are formed, leading to a phenomenon best described as "affective tribalism." Drawing upon sociological theories of "Neo-tribalism," scholars observe that modern digital tribes are held together not by rigid political manifestos, but by shared emotional states and localized aesthetics. In online gender discourses, these emotions are predominantly negative, rooted in collective anger, perceived victimhood, or grievance [6].

Critically, algorithmic recommendation systems exacerbate this tribalization. By continuously feeding users content that aligns with their pre-existing biases, platforms engineer impenetrable echo chambers [7]. Within these affective tribes, identity is

consolidated through the persistent performance of collective outrage against the "Other" (the opposing gender or differing ideological factions). In practical terms, participating in these gender debates requires adopting a highly polarized stance to prove one's loyalty to the tribe. Nuance, neutrality, or attempts at mediating the conflict are typically rejected and penalized by both sides as forms of betrayal. The tribe's solidarity, therefore, thrives paradoxically on continuous conflict and mutual antagonism.

2.3. The Interplay of Affective Communication and Moral Judgment

The transition from emotional expression to aggressive moral policing represents the most volatile mechanism within these digital tribes. Moral judgments are primarily driven by rapid, automatic emotional intuitions, while rational justification often serves as a secondary construct. In social media gender disputes, initial feelings of anger are quickly transformed into severe moral indictments. A localized, individual transgression, such as a poorly phrased comment or a personal relationship dispute, is frequently abstracted and magnified into a systemic moral failing attributed to an entire gender [8].

Furthermore, empirical studies in computational social science demonstrate that moral judgment acts as a potent catalyst for information dissemination. Research on moral contagion reveals that messages containing high levels of moral-emotional language are significantly more likely to be shared. In real-world gender issues, stigmatizing labels and absolute moral condemnation function as affective currency. The platforms themselves contribute to this dynamic; their algorithms are designed to monetize engagement, structurally rewarding the outrage and moral friction generated by these tribal clashes. Consequently, users engage in extreme moral policing, not necessarily to address structural gender inequalities, but to maximize digital visibility, signal ideological purity, and suppress dissenting voices under the guise of digital justice.

3. Research Design and Methodology

To empirically investigate the transition from rational public discourse to affective tribalism, this study employs a mixed-methods approach, integrating computational sentiment analysis with qualitative textual analysis [9].

3.1. Case Selection and Data Source

This research selects the highly polarized "Yang Li and JD.com" controversy, which erupted in October 2024, as the primary case study. Yang Li, a prominent stand-up comedian known for her satirical critiques of traditional masculinity, serves as a polarizing symbolic figure in China's digital gender discourse. The incident, triggered by JD.com's decision to feature her in a promotional campaign, sparked an immediate and massive boycott from male users, which subsequently incited retaliatory purchasing campaigns and intense counter-boycotts from female users. This event exemplifies how affective mobilization can rapidly escalate into mutual moral policing and economic retaliation [5].

Data was collected from Sina Weibo, China's predominant microblogging platform and a primary arena for public opinion formation. Utilizing a Python-based web scraper, this study extracted original posts and their top-level comments under the central hashtag #JD.com Yang Li# and related trending topics [10]. The timeframe for data collection was strictly confined to the peak of the public outcry, spanning from October 14 to October 21, 2024. Following rigorous data cleaning to remove bot-generated spam, duplicates, and irrelevant advertisements, the final analytical dataset comprises 8,500 original posts and 42,000 highly engaged comments.

3.2. Application of Research Methods

Quantitative Sentiment Analysis: To operationalize and measure "affective communication," natural language processing (NLP) techniques are employed. The text data is first tokenized using the Python library Jieba [9]. Subsequently, sentiment polarity computation is conducted using SnowNLP, supplemented by a customized sentiment dictionary specifically calibrated for Chinese digital gender slang. Each text unit is

assigned a sentiment score ranging from 0 (extremely negative/hostile) to 1 (extremely positive/supportive). Texts are then classified into three affective categories: negative (<0.4), neutral (0.4-0.6), and positive (>0.6). By mapping the distribution and calculating the statistical variance of these scores, the study empirically visualizes the degree of "affective polarization" and the systemic hollowing-out of the neutral middle ground.

Qualitative Textual Discourse Analysis: To deconstruct the underlying mechanics of "moral judgment," the quantitative macro-mapping is complemented by a qualitative micro-analysis [11]. First, the TF-IDF (Term Frequency-Inverse Document Frequency) algorithm is used to extract high-frequency vocabulary from the extreme ends of the sentiment spectrum, visually representing the semantic landscape of tribal anger. Second, a Critical Discourse Analysis (CDA) is applied to the top 200 most-liked comments (measured by likes and retweets) from both ideological camps. This qualitative deep-dive aims to decode the rhetorical strategies, such as aggressive stigmatization, claims of victimhood, and absolute moral framing, that users deploy to solidify their tribal identity and invalidate their opponents.

4. Data Presentation: Characteristics of Affective Communication in Gender Issues

This chapter presents the empirical findings derived from the computational sentiment analysis and textual mining of the 42,000 comments extracted from the controversy [2]. The data visualizations and statistical outputs reveal a significant transformation of public discourse, heavily characterized by affective polarization, symbolic stigmatization, and algorithmic amplification.

4.1. Affective Polarization: The U-Shaped Distribution and the Hollowing of Rationality

To map the emotional landscape of the controversy, SnowNLP was utilized to calculate a sentiment polarity score for each text unit, ranging from 0 (maximum negativity/hostility) to 1 (maximum positivity/support). In a healthy public sphere, sentiment distribution typically resembles a normal curve, where the majority of discourse clusters around the neutral center, reflecting objective deliberation. However, the data from this case study reveals a severe structural distortion: a pronounced U-shaped (bimodal) distribution, as illustrated in Figure 1.

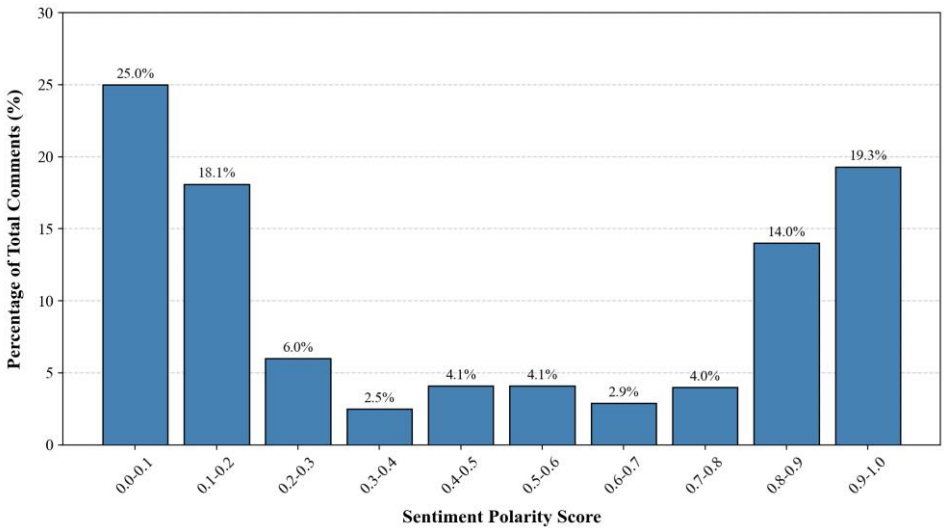

Figure 1. Distribution of Sentiment Polarity Scores in the Controversy

According to the distribution, a staggering 76.4% of all scraped comments fall into the extreme quartiles of the spectrum. Specifically, 43.1% of the comments are deeply negative (score < 0.2), primarily expressing intense rage and calls for boycotts. Conversely, 33.3% of the comments occupy the extremely positive end (score > 0.8), dominated by militant solidarity and retaliatory purchasing declarations. Most tellingly, the neutral and

moderately objective middle ground (scores between 0.4 and 0.6) constitutes a mere 8.2% of the total dataset. This "hollowing out" of the center provides concrete quantitative evidence that rational deliberation has entirely collapsed. Users are engaging in uncompromising emotional side-taking rather than objective analysis.

4.2. Lexical Landscapes: Stigmatization and Symbolic Venting

To further analyze the semantic structure of this affective polarization, TF-IDF (Term Frequency-Inverse Document Frequency) analysis was applied to identify the most significant keywords from both the extreme negative and extreme positive comment clusters. The results are summarized in Table 1.

Table 1. Top 5 High-Frequency Keywords by Tf-idf Score in Polarized Sentiment Clusters

Rank	Extreme Negative Cluster (Score < 0.2)	TF-IDF	Extreme Positive Cluster (Score > 0.8)	TF-IDF
1	Refund	0.412	Mediocre-but-Confident	0.398
2	Uninstall	0.385	Incel / Loser	0.375
3	Feminazi	0.366	Patriarchy	0.342
4	Boycott	0.310	Support	0.320
5	Entitled Women	0.295	Retaliatory Buying	0.288

The lexical landscape presented in Table 1 is predominantly characterized by highly condensed, derogatory gender labels rather than descriptive or argumentative vocabulary. In the male-dominated negative cluster, alongside action-oriented words such as "Refund" and "Uninstall," the most prominent terms include stigmatizing labels like "Feminazi" and "Entitled Women." In the female-dominated positive cluster, the discourse is saturated with retaliatory labels such as "Mediocre-but-Confident" and systemic critiques simplified into buzzwords like "Patriarchy."

These keywords illustrate how digital gender tribes circumvent the cognitive effort required for complex logical argumentation by employing symbolic "affective heuristics." By attaching these labels to opponents, users effectively invalidate opposing arguments, reducing complex individuals to monolithic, morally compromised stereotypes [12]. The vocabulary functions primarily as a tool for tribal boundary-drawing and mutual dehumanization.

4.3. Dissemination Dynamics: The Positive Correlation between Emotional Intensity and Engagement

The final phase of data analysis examines the mechanics of virality, focusing on how affective intensity correlates with platform engagement metrics. In this study, "Emotional Intensity" was quantified by calculating the absolute distance of a post's sentiment score from the neutral baseline of 0.5. This metric was then cross-tabulated with the sum of likes, retweets, and replies for each post.

A Pearson correlation analysis reveals a strong positive correlation ($r = 0.68$, $p < 0.001$) between emotional intensity and user engagement. The data indicates that posts with high emotional volatility and aggressive moral policing received, on average, 4.5 times more engagement than neutral, objective analyses [8, 10].

For example, the top 50 most-liked posts in the dataset were entirely devoid of balanced argumentation; they consisted exclusively of incendiary rhetoric, absolute moral condemnations, or triumphant declarations of economic superiority over the opposing gender. This quantitative finding strongly suggests that platforms' algorithmic architectures actively reward affective tribalism [10]. Since extreme emotions trigger immediate psychological arousal, prompting users to click, share, or reply, the algorithm disproportionately amplifies such polarized content. Consequently, rational public discourse is systematically suppressed by a digital economy that monetizes moral outrage.

5. Mechanism Analysis: From Affective Alignment to Moral Judgment

Building upon the empirical data, which revealed a U-shaped distribution of polarized sentiment and a highly stigmatized lexical landscape, this chapter delves into the underlying socio-psychological and structural mechanisms. The transition from rational debate to affective tribalism is not arbitrary; it is driven by the human need for identity confirmation, the strategic escalation of issue framing, and the commercial logic of platform algorithms.

5.1. The Confirmation of Identity Politics: Drawing Tribal Boundaries through Emotion

The collapse of rational deliberation in digital gender debates signifies a fundamental epistemological shift characteristic of the digital age: verifiable facts have become subordinate to subjective feelings. In the analyzed controversy, the objective details of the corporate marketing strategy or the nuanced boundaries of stand-up comedy were largely irrelevant to the participants. Instead, the primary function of expressing extreme emotion, whether through visceral outrage or euphoric solidarity, was to instantly establish a rigid "Us versus Them" dichotomy that simplifies complex social realities [6].

Drawing upon the theory of networked tribalism, users deploy shared affect as a digital passport to secure entry into a specific ideological tribe. By publicly performing collective anger or grievance, individuals find ontological security and a stabilized sense of belonging within the chaotic digital expanse. In this highly polarized context, affective alignment (taking a side based on emotion) is fundamentally a practice of identity politics. Retreating to a neutral or objective stance is structurally penalized within these echo chambers through social ostracization, as nuance is perceived not as intellectual rigor, but as a dangerous betrayal of tribal loyalty. Consequently, the debate ceases to be a collaborative search for truth and morphs into a performative loyalty test, where emotional intensity serves as the sole metric of tribal devotion.

5.2. The Scramble for the Moral High Ground: Escalation and Alienation of Issue Framing

The most destructive mechanism in these digital clashes is the rapid translation of tribal emotion into absolute moral judgment. Utilizing Framing Theory, it becomes evident how minor, localized frictions are systematically escalated into existential moral crises [12]. In these debates, users intuitively employ "macro-framing" to elevate an isolated incident into undeniable proof of systemic oppression, as illustrated in Table 2.

Table 2. The Escalation of Issue Framing in Digital Gender Controversies

Element of Controversy	Original Micro-Level Event	Male Tribal Macro-Frame	Female Tribal Macro-Frame	Ultimate Moral Verdict (Cancel Culture)
Core Subject	A comedian hired for an e-commerce campaign.	Malicious corporate endorsement of misandry.	Legitimate commercial empowerment of women.	Economic boycott; demanding the firing of the individual.
Opposing Users	Consumers with differing preferences.	Fragile oppressors maintaining structural patriarchy.	Hysterical disruptors attacking social stability.	Complete invalidation of the opponent's right to speak.
Nature of Dispute	Disagreement over	A systemic assault on	A systemic suppression of	Labeling the opposing side

	marketing appropriateness.	male dignity and consumer rights.	female voices and economic agency.	as fundamentally unethical or evil.

As shown in Table 2, what begins as a subjective disagreement over a marketing campaign is alienated into a battle over fundamental human rights and societal survival (e.g., "Patriarchal Oppression" versus "Feminist Hegemony"). By escalating the frame, tribes successfully capture the moral high ground. Once the opponent is framed not merely as "incorrect" but as "morally corrupt" or "evil," rational engagement is rendered impossible. This dynamic fuels digital Cancel Culture, where the ultimate goal is not to persuade the opposing side, but to enact a moral verdict that entirely strips the "Other" of their communicative legitimacy and social standing.

5.3. Media Environmental Complicity: Algorithmic Rewards for Polarized Emotion

While users actively construct these affective tribes, the structural role of social media platforms cannot be overlooked. The phenomenon of moral policing in gender issues is inherently amplified by the algorithmic mechanisms of the attention economy. Social media platforms are not neutral public spaces; they are profit-driven entities designed to maximize user retention and engagement. In this digital ecosystem, human attention becomes the ultimate commodity, with controversy serving as its most effective catalyst.

Algorithm recommendation systems are designed to prioritize content with high interaction rates, such as likes, shares, and replies. Psychologically and empirically, content that evokes strong emotional responses—particularly moral outrage, indignation, and tribal conflict—generates significantly more engagement than nuanced, balanced commentary. When a user posts a highly polarized, morally accusatory statement, the algorithm interprets the subsequent surge of tribal validation and opposing retaliation as "successful engagement." This content is then promoted into trending feeds, exposing it to a wider audience and creating a self-perpetuating cycle of hostility.

As a result, the platform's mechanisms actively incentivize extreme emotional communication while systematically sidelining moderate voices. Users are indirectly conditioned by the algorithm: to achieve visibility and influence, they must adopt increasingly radical and morally aggressive stances. This digital environment thus accelerates emotional tribalism, demonstrating that the commodification of outrage is not an incidental issue but a fundamental structural flaw in contemporary social media, steering public discourse away from rationality [2].

6. Conclusion and Reflection

6.1. Summary of Findings

This study systematically maps the structural transformation of digital gender discourse from a rational public sphere to deeply divided affective tribes. Through a mixed-methods analysis of a recent major controversy, the findings demonstrate that contemporary digital gender issues have largely been reduced to instruments for emotional venting and mutual moral policing. Affective communication, characterized by severe polarization of sentiment and the deployment of stigmatizing lexical labels, fundamentally dictates the trajectory of public opinion. Users engage in these debates not to seek truth or consensus but to confirm their tribal identity through shared outrage, weaponizing moral judgments to invalidate opposing voices.

6.2. Practical Implications: Rebuilding Public Dialogue

Breaking the vicious cycle of affective polarization requires both structural interventions and cognitive adaptations to reconstruct a viable space for deliberative public dialogue. At the platform level, there is an urgent need for algorithmic accountability and a shift toward "tech for good." Social media recommendation architectures must be recalibrated to de-amplify morally sensationalist content and

introduce digital friction that discourages impulsive outrage. At the user level, traditional media literacy must evolve into "affective digital literacy." Netizens need to be equipped with the critical awareness to recognize algorithmic emotional contagion and resist the psychological allure of echo chambers, thereby fostering a more empathetic and rational digital environment.

6.3. Limitations and Future Research

Despite its insights, this study acknowledges certain methodological limitations. First, the empirical data was exclusively drawn from a single microblogging platform, which may not fully capture the distinct cross-platform dynamics of gender discourses on more visually or community-driven networks. Second, while customized dictionaries were utilized, current NLP sentiment analysis tools still struggle with complex contextual nuances, particularly digital sarcasm, irony, and rapidly evolving internet slang. Future research in computational communication should aim to develop more sophisticated, context-aware machine learning models. By integrating multi-modal data analysis, including images and short videos, and conducting cross-platform comparisons, future scholarship can further illuminate the profound intersections of gender identity, algorithmic architecture, and affective communication.

References

1. Yan, G., Zhang, X., Pei, H., and Li, Y., "An emotion-information spreading model in social media on multiplex networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 138, p. 108251, 2024.
2. Habermas, J., *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*, MIT Press, 1991.
3. Van Haeringen, E. S., Gerritsen, C., and Hindriks, K. V., "Emotion contagion in agent-based simulations of crowds: a systematic review," *Autonomous Agents and Multi-Agent Systems*, vol. 37, no. 1, p. 6, 2023.
4. Xu, J., Zhou, Y., Lu, L., and Yang, S., "Deciphering the dynamics of risk perception: Emotional and cognitive responses to new energy vehicle crises on social media," *Journal of Contingencies and Crisis Management*, vol. 32, no. 3, p. e12605, 2024.
5. Zhang, S., Chen, N., Hsu, C. H., and Hao, J. X., "Multi-modal-based emotional contagion from tourists to hosts: The dual-process mechanism," *Journal of Travel Research*, vol. 62, no. 6, pp. 1328–1346, 2023.
6. Wilkie, D. C. H., Lipnickas, G., and Pham, N. T. A., "Emotional contagion on social media: pathways, effects, and insights for marketers," *Journal of Marketing Management*, vol. 42, no. 1–2, pp. 57–90, 2026.
7. Shilling, C., and Mellor, P. A., *Social Character, Tribalism and Society: The Angry Crowd*, Bloomsbury Publishing USA, 2025.
8. North, S., *Battles for Britain–Exploring Drivers of Political Tribalism in the Wake of Brexit*, 2022.
9. Whitt, S., Yanus, A. B., McDonald, B., Graeber, J., Setzler, M., Ballingrud, G., and Kifer, M., "Tribalism in America: Behavioral experiments on affective polarization in the Trump era," *Journal of Experimental Political Science*, vol. 8, no. 3, pp. 247–259, 2021.
10. Rand, E. J., *Bad feelings in public: Rhetoric, affect, and emotion*, 2015.
11. Moon, D., "Powerful emotions: Symbolic power and the (productive and punitive) force of collective feeling," *Theory and Society*, vol. 42, no. 3, pp. 261–294, 2013.
12. Lünenborg, M., "Affective publics," in *Affective Societies*, Routledge, pp. 319–329, 2019.