Psychology & Marketing

WILEY

RESEARCH ARTICLE OPEN ACCESS

Don't You Know That You're Toxic? How Influencer-Driven Misinformation Fuels Online Toxicity

Giandomenico Di Domenico[1] | Federico Mangió[2] | Denitsa Dineva[1]

[1]Cardiff Business School, Cardiff University, Cardiff, UK | [2]Department of Management, University of Bergamo, Bergamo (BG), Italy

Correspondence: Giandomenico Di Domenico (didomenicog@cardiff.ac.uk)

Received: 21 August 2025 | **Revised:** 7 January 2026 | **Accepted:** 9 January 2026

Keywords: audience polarization | digital misinformation | social media influencers | symbolic brand attacks | toxic online discourse

ABSTRACT

Research on misinformation has focused on message content and cognitive bias, overlooking how source type shapes toxic engagement. This study addresses that gap by showing that influencer-driven misinformation does not merely increase toxicity: it reconfigures its nature and persistence through relational and social influence mechanisms. Drawing on Source Credibility, Parasocial Interaction, and Social Influence theories, we analyse 101 brand-related misinformation posts (48,821 comments) across major platforms using a mixed-method design combining automated toxicity detection, topic modeling, and thematic analysis. Results reveal that influencers amplify toxicity under high engagement, sociopolitical salience, and low pseudonymity conditions, producing distinct patterns such as flame-bait firestorms and toxic debunking. We identify two influencer-specific mechanisms: brand-related misinformation legitimation and community enmeshment, that sustain toxic echo chambers by converting credibility and parasocial bonds into collective antagonism. These findings advance marketing theory by reframing toxicity as a source-amplified, relational phenomenon, and inform ecosystem-level interventions structured around publishers, platforms, and people to mitigate influencer-driven harm.

1 | Introduction

In early 2025, several social media influencers on TikTok shared viral videos alleging that luxury brands such as Hermès, Louis Vuitton, and Chanel secretly manufacture their goods in Chinese factories while falsely marketing them as "Made in France" or "Made in Italy." The influencers presented their claims as exposés of industry deceit, despite offering no verifiable evidence to support them (Hall 2025). The videos amassed millions of views and stimulated widespread debate among users concerning authenticity, ethical conduct, and transparency within the luxury sector, positioning the implicated brands at the center of online criticism and misinformation.

This case highlights a growing paradox in influencer culture. Brands increasingly rely on social media influencers (SMIs) to reach and engage with target audiences, with the market reaching a record of 24 billion U.S. dollars in 2024 (Statista 2023). Despite the positive impact of SMIs on marketing outcomes (Gurrieri et al. 2023; Leung et al. 2022), their prominence also introduces new risks, particularly when controversial or misleading content sparks toxic reactions directed at brands. While recent studies have started to examine the role of SMIs in spreading false or inflammatory material (Ekinci et al. 2025; Harff et al. 2022; Stewart et al. 2023, the mechanisms linking misinformation and online toxicity remain underexplored.

Understanding whether toxicity unfolds differently when misinformation originates from regular users versus SMIs is vital, given the distinct levels of influence, credibility, and audience engagement they command, with influencers strategically using platform features and personal branding to amplify credibility and engagement (Gurrieri et al. 2023; Scholz 2021). The financial incentives and engagement-driven nature of social media further increase the likelihood that SMIs contribute to misinformation and toxicity: content forms that are, ironically, among the most rewarding in terms of visibility and reach

(Avalle et al. 2024; Cinelli et al. 2021; Vosoughi et al. 2018). This dynamic is particularly salient in the context of brand-related misinformation, where influencer marketing and symbolic consumption intersect. Influencer collaborations represent a major source of marketing revenue, while brands embody symbolic, identity, and experiential value that extend beyond their functional attributes (Cova and Dalli 2009; Holt and Cameron 2010). Consequently, misinformation involving brands not only distorts factual information but also threatens consumers' identity-related meanings, often intensifying emotional and polarized reactions online (Visentin et al. 2019).

While prior research has primarily examined the *content* of misinformation (Di Domenico et al. 2022; Pennycook and Rand 2021; Johar 2022), less attention has been paid to its *source*. Social psychology research indicates that engagement with misinformation is often driven more by social identity and influence processes than by accuracy motives (Van Bavel et al. 2024). Within this context, influencers occupy a unique position: their perceived expertise and authenticity (Ohanian 1990) encourage followers to process information with trust and emotional involvement (Bovet and Makse 2019). Through parasocial relationships (Horton and Richard Wohl 1956), audiences often experience a sense of intimacy and reciprocity that makes influencers' opinions feel personally meaningful (Lou 2022). When these figures share misleading or controversial claims, followers may internalize and defend such content as part of their identity expression. These dynamics reflect classic social influence mechanisms (Kelman 1958) and suggest that misinformation spread by influencers may evoke more intense, belief-consistent, and polarized reactions than similar content originating from regular users, who lack comparable credibility or relational depth. Over time, these influence processes can sustain and amplify a toxic spiral, as misinformation and emotionally charged engagement become mutually reinforcing within online communities.

To examine this dynamic, we adopted an empirics-first approach (Golder et al. 2023), compiling a multiplatform dataset of brand-related misinformation posts and associated user comments spanning 47 brands across nine industries over a 3-year period (2020–2023). We then implemented a sequential mixed-method design that integrated exploratory and explanatory techniques, including top-down automated textual analysis for toxicity detection (Humphreys and Wang 2018), bottom-up topic modeling, and theory-building thematic analysis.

Our findings show that influencers spreading brand-related misinformation generate markedly higher toxicity than regular users, exhibiting distinct behavioral patterns. Five categories of toxicity emerge: anti-brand reactions, consumer-to-consumer conflict, flame-bait firestorms, toxic debunking, and trolling or flaming. Regular users elicit more heterogeneous out-group hostility and corrective aggression, whereas influencers primarily trigger flame-bait firestorms that consolidate echo chambers. Two mechanisms underpin this process: *brand-related misinformation legitimation*, through which influencers amplify and defend misinformation, and *community enmeshment*, through which they deepen follower identification and emotional alignment, sustaining toxic engagement over time.

This study advances theory at the intersection of misinformation, influencer marketing, and online incivility in three ways. First, it expands prior work on harmful influencer practices (Bahar and Hasan 2024; Ekinci et al. 2025; Karagür et al. 2022), showing that misinformation disseminated by SMIs generates more frequent and qualitatively distinct toxicity than identical content from regular users. Rather than locating harm in message content alone, the findings theorize toxicity as a source-amplified outcome, whereby influencer credibility legitimizes misinformation and lowers normative barriers to incivility (Di Domenico et al. 2022). In this way, credibility shifts from a persuasive asset to an infrastructural mechanism through which toxicity is normalized. Second, the study reconceptualises online toxicity as a relational and performative outcome of influencer-led misinformation. It extends existing typologies (Fombelle et al. 2020; Martel et al. 2024) by distinguishing established forms of incivility from emergent patterns—such as flame-bait firestorms and toxic debunking—that arise specifically in influencer contexts where authority and audience alignment shape the targets and moral framing of aggression. Third, the findings challenge cognition-centric accounts of misinformation processing by showing that compliance, identification, and internalization operate as collective alignment mechanisms (Kelman 1958; Van Bavel et al. 2024). Misinformation persistence and toxicity thus emerge from social reinforcement and identity dynamics rather than informational deficits alone.

Managerially, this study provides actionable guidance for mitigating influencer-driven toxicity by focusing on ecosystem-level interventions. Specifically, we outline strategies for early detection and containment through platform-level monitoring, brand response protocols, and community-oriented approaches that address relational dynamics and prevent escalation.

2 | The Phenomenon: The Spread and Consequences of Brand-Related Misinformation

2.1 | Misinformation Spread

The spread of misinformation on social media poses a major challenge for brands and consumers. False or misleading content shapes attitudes, distorts brand narratives, and provokes toxic exchanges between users (Guldemond et al. 2022). Extensive research shows that misinformation spreads faster, farther, and more broadly than factual information, fueled by emotional content, novelty, and algorithmic amplification (Vosoughi et al. 2018; Del Vicario et al. 2016). This diffusion is reinforced by social endorsement cues, repetition, and identity-consistent sharing, which strengthen perceived accuracy and encourage re-posting (Pennycook and Rand 2021; Brady et al. 2023).

While misinformation has been extensively studied in political and health domains (Di Domenico et al. 2022; Vosoughi et al. 2018), brand-related misinformation represents a distinct context shaped by the symbolic and relational nature of brands. Brands differ from political or institutional targets because consumers often experience them as extensions of personal identity, moral stance, and group belonging (Cova and Dalli 2009; Gensler et al. 2013). This participatory and commercial dimension creates a sense of symbolic ownership, meaning that false or misleading information about a brand can evoke not only opinion disagreement but feelings of personal

affront and collective defence. As consumers emotionally invest in brands, they can become divided into vocal supporters and critics, particularly in reactive social media environments (Ammann et al. 2025).

Misinformation campaigns exploit this symbolic attachment. By distorting brand meanings, they provoke emotional responses, reinforce in-group bonds, and fuel engagement: dynamics intensified by algorithms that reward controversy (Bahar and Hasan 2024). In algorithmically governed environments, engagement metrics serve as implicit credibility signals, enabling misinformation to gain visibility through likes, shares, and comments regardless of its truth value (Cinelli et al. 2021). Such dynamics transform social platforms into self-reinforcing ecosystems where emotionally charged falsehoods thrive.

For a better understanding of how misinformation spreads, it is equally important to consider who produces and circulates it first (Di Domenico et al. 2021). Misinformation often originates from elite or high-visibility sources such as politicians or media figures (Grinberg et al. 2019) and is later amplified by regular users (Aral 2020) or automated networks of bots (Shao et al. 2018). In social media ecosystems, these visible actors act as credibility gateways, legitimizing misleading narratives through status (Bovet and Makse 2019) and perceived expertise (Di Domenico et al. 2022). While ordinary users can perpetuate false information, SMIs occupy a distinct communicative position, as they blur the boundary between peer and celebrity. Their reach, visibility, and perceived authenticity make them powerful intermediaries between brands and audiences (Shehzala et al. 2024). When misinformation originates from an influencer rather than a regular user, it can provoke stronger emotional reactions and higher levels of toxic commentary because audiences perceive the influencer as more credible and personally relevant.

To contextualize these dynamics, Table 1 summarizes prior research examining influencers' role in misinformation, moral manipulation, and toxicity. The table outlines each study's focus, theoretical lens, and contribution to understanding influencer behavior, highlighting how the present study extends this body of work.

2.2 | Toxicity as a Consequence

Online toxicity refers to hostile, aggressive, or derogatory communication that disrupts constructive interaction. Prior research links toxic behavior to factors such as anonymity, deindividuation, emotional contagion, and moral outrage (Brady et al. 2023; Suler 2004). However, little attention has been paid to how the credibility or relational position of a message source shapes such responses. Toxicity is increasingly recognized as a by-product of misinformation, particularly when it targets emotionally charged or symbolic entities such as brands (Papakyriakopoulos and Goodman 2022). In these contexts, false or misleading claims not only distort perception but also provoke strong affective reactions and reputational harm (Mills and Robson 2020).

Building on this view, this study focuses on *interactive* toxicity: reciprocal, often public exchanges that amplify visibility and emotional contagion. Such interactions include trolling, harassment, and coordinated attacks that thrive on engagement mechanics such as likes, shares, and algorithmic promotion (Walker et al. 2019; Vogels 2021; Dineva 2023). Unlike isolated abusive comments, interactive toxicity evolves relationally: it circulates within communities, reinforcing group boundaries and escalating misinformation through social validation (Colliander 2019). Understanding these interactive patterns is crucial to explain how influencer-driven misinformation transforms digital conversations into sustained toxic engagement.

Research has identified multiple forms of online toxicity and their linguistic characteristics, including insults, swearing, and aggression (Dineva et al. 2020). These behaviors manifest both interpersonally and toward brands. Interpersonal toxicity involves exchanges such as trolling (Golf-Papez and Veer 2022), flaming (Cho and Kwon 2015), harassment (Vogels 2021), and consumer-to-consumer conflict (Dineva et al. 2020), which often arise within brand communities or between rival consumer groups (Ewing et al. 2013; Husemann et al. 2015; Dineva and Daunt 2023). Brand-directed toxicity, by contrast, targets the organization itself, typically in response to perceived misconduct, unsatisfactory service, or competition, and takes the form of malicious word-of-mouth (Hornik et al. 2019), firestorms (Herhausen et al. 2019), anti-brand activism (Romani et al. 2015), or brand trolling (Dineva and Breitsohl 2022).

While this body of work has advanced understanding of how toxicity emerges and spreads, it largely centers on user-generated behavior rather than the influence of message sources. Studies have begun to consider user typologies and patterns that drive hostility (Bacile et al. 2025; Golf-Papez and Veer 2022; Kim et al. 2021), yet the role of social media influencers in triggering or amplifying toxicity remains underexplored. Influencers differ from ordinary users in their visibility, credibility, and ability to shape discourse through parasocial relationships. Some research links them to polarization and conflict (Koorank Beheshti et al. 2023) or to the use of controversy and aggression as self-branding strategies (Abidin 2019).

3 | Theoretical Foundations

Although research on misinformation and online toxicity has advanced (Di Domenico et al. 2022; Cinelli et al. 2021), marketing scholarship still lacks an integrated framework connecting cognitive, relational, and collective dimensions of influence. To address this gap, we draw on source credibility theory (Hovland and Weiss 1951; Ohanian 1990), parasocial interaction theory (Horton and Richard Wohl 1956), and social influence theory (Kelman 1958) to explain how influencers' perceived expertise, emotional closeness, and group-based reinforcement transform persuasive communication into toxic, polarized exchanges. These lenses guide our three-stage investigation: Study 1 documents the phenomenon using large-scale automated text analysis, Study 2 explains its drivers through topic modeling and regression, and Study 3 uncovers sustaining mechanisms via thematic analysis. This synthesis responds to calls for greater theoretical precision (Golder et al. 2023) by showing how credibility, attachment, and social alignment jointly shape reactions to influencer-driven misinformation.

Given the study's exploratory orientation, the research proceeds through theory-driven research questions (RQs) rather than

TABLE 1 | Summary of prior studies on influencer effects and positioning of the present research.

References	Context	Key constructs	Theoretical lens	Main findings	Relevance
Abidin (2019)	Influencer behavior in conflict situations	Self-branding, moral positioning	Critical discourse/ identity theory	Influencers may engage in bullying or retaliatory behaviors for visibility	Introduces influencer toxicity and moral manipulation as a visibility strategy
Barari (2023)	Online political influencers	Echo chambers, misinformation	Moral foundations/ identity theory	Influencers act as amplifiers of outrage through moral framing	Demonstrates how influencers' framing power fuels emotional toxicity, extends to brand contexts in our work
Koorank Beheshti et al. (2023)	Influencer-driven polarization	Hate speech, community identity	Social identity theory	Influencers amplify moral and ideological divides online	Directly supports our view of influencer-driven toxicity as relational and group-based
Breves et al. (2019)	Instagram influencer endorsements	Credibility, trust, persuasion	Source credibility theory	Influencer expertise and trustworthiness increase persuasion effectiveness	Highlights how perceived credibility drives positive engagement: Our study tests when it turns negative under misinformation
Di Domenico et al. (2022)	Health misinformation (COVID-19)	Misinformation, credibility, influence	Legitimacy theory	Influencers disseminated misleading health information due to perceived expertise	Parallels the mechanism of influencer credibility legitimizing misinformation in brand contexts
Han and Balabanis (2024)	Influencer credibility and brand attitudes	Source credibility, consumer trust	Source credibility theory	Influencer credibility enhances brand evaluations	Establishes the credibility–trust mechanism that, under misinformation, may lead to expectancy violations and toxicity
Hughes et al. (2019)	Consumer–influencer relationships	Authenticity, engagement, parasocial closeness	Parasocial interaction theory	Authentic influencers foster trust and long-term engagement	Provides baseline for understanding relational closeness that can also fuel toxic defence when misinformation occurs
Mulcahy et al. (2024)	Health misinformation	Perceived deception, parasocial interaction, and sharing intentions	Social influence and cognitive appraisal theorie	When influencers share misinformation, virality metrics determine lower deception and higher parasocial relationship and sharing intentions	One of the few investigations on consumers' responses to influencer-driven misinformation
Shehzala et al. (2024)	Influencer marketing and self-perception	Homophily, authenticity, self-acceptance	Source attractiveness model	Perceived similarity and mindfulness increase positive influencer outcomes	Informs our understanding of audience–influencer alignment and emotional investment

(Continues)

TABLE 1 | (Continued)

References	Context	Key constructs	Theoretical lens	Main findings	Relevance
Valsesia et al. (2020)	Influencer self-presentation	Self-congruence, authenticity	Self-congruity theory	Congruent self-presentation enhances influencer appeal	Connects identity expression with influencer effectiveness; our study explores the identity backlash side
Wallace and Buil (2025)	Disinformation and opinion leadership	Personality traits, perceived credibility	Personality and persuasion	Dark triad traits predict disinformation sharing	Highlights individual-level antecedents; our study focuses on audience-level reactions to influencer misinformation
This study	Brand-related misinformation across six social media platforms (2020–2023)	Source type, toxicity, misinformation legitimation, community enmeshment	Source credibility theory, parasocial interaction theory, and social influence theory	Shows that influencer-driven misinformation generates more intense and relationally embedded toxicity than user-driven misinformation	Extends prior research by integrating credibility and relational theories to explain *how* and *why* influencer misinformation amplifies toxic engagement in brand contexts

formal hypotheses. This approach aligns with established guidance on conceptual and mixed-method inquiry in marketing, which recommends research questions as a means of articulating theoretically grounded expectations while preserving analytical flexibility (Edmondson and McManus 2007).

3.1 | Source Credibility and the Magnitude of Toxic Responses to Misinformation

Influencers actively contribute to misinformation spread through strategic credibility signaling and self-presentation (Bovet and Makse 2019; Bahar and Hasan 2024). The COVID-19 pandemic emphasized this issue: a small number of highly visible influencers were responsible for a disproportionate share of vaccine-related misinformation (CCDH 2021). Even before the pandemic, influencers had used perceived expertise to promote misleading claims in areas such as health and wellness (Di Domenico et al. 2022).

Source credibility theory (Hovland and Weiss 1951; Ohanian 1990) posits that a communicator's perceived expertise, trustworthiness, and attractiveness shape how audiences interpret messages. In digital contexts, influencers' credibility is a primary driver of audience attachment and engagement (Shehzala et al. 2024). Perceived expertise and authenticity not only enhance persuasive effectiveness but also encourage followers to interact more frequently, emotionally, and publicly with influencer content (Audrezet et al. 2020). High credibility thus functions as both a persuasive cue and a relational amplifier, strengthening followers' psychological investment and visibility within the influencer's community (Yuan and Lou 2020).

This intensified engagement can, however, magnify the emotional consequences of misinformation. When trusted influencers share misleading or controversial content, their credibility heightens followers' motivated responses, from loyal defence and mimicry to moral outrage and toxic confrontation (Abidin 2019; Campbell and Farrell 2020; Koorank Beheshti et al. 2023). In such contexts, credibility transforms from a mechanism of persuasion into one of polarization, as audiences' strong identification with credible influencers fuels algorithmically amplified toxic exchanges (Garibay et al. 2019; Mulcahy et al. 2024).

Communication studies consistently show that source characteristics shape persuasion and emotional response (Hovland and Weiss 1951; Metzger et al. 2021), yet marketing research has seldom examined how different message sources activate toxicity in brand-related misinformation contexts. This context is relevant as brand-related misinformation offers influencers symbolic and engagement value. Engaging with brands (positively or negatively) enables influencers to amplify their visibility and relevance. Because social media algorithms reward polarizing and emotionally charged content, misinformation becomes both a visibility tactic and a discursive resource (Koorank Beheshti et al. 2023; Garibay et al. 2019). In these environments, SMIs occupy a hybrid position: they are simultaneously peers, endorsers, and opinion leaders (Audrezet et al. 2020; Han and Balabanis 2024). Their perceived credibility and relational closeness grant them persuasive power that may not simply persuade but also polarize (Abidin 2019; Koorank Beheshti et al. 2023).

Collectively, these insights suggest that credibility may not merely facilitate persuasion but also intensify toxicity when misinformation originates from influential rather than ordinary users. Yet, empirical research has not systematically compared how misinformation from these different *sources* shapes audience reactions. This leads to our first research question.

RQ1. *How does the source of brand-related misinformation (influencers vs. regular users) influence the intensity and nature of toxic audience responses on social media?*

3.2 | Parasocial Relationships and the Type of Toxic Responses to Misinformation

Parasocial interaction theory (Horton and Richard Wohl 1956) explains how audiences form one-sided, affective relationships with media figures. These imagined yet emotionally significant relationships create a sense of intimacy, familiarity, and reciprocity, even in the absence of direct interaction. In digital environments, influencers have become paradigmatic examples of this process. Through consistent self-disclosure, interactive communication, and everyday visibility, they foster a perception of authentic connection that encourages followers to engage as if participating in a mutual relationship (Lou 2022).

Influencers further occupy a central position in marketing communication, shaping attitudes, behaviors, and brand perceptions through their perceived authenticity, accessibility, and visibility (Han and Balabanis 2024). Their effectiveness depends on cultivating parasocial ties that foster emotional closeness and trust (Reinikainen et al. 2020; Thomas et al. 2024), embedding them within their audiences' online communities (Mardon et al. 2018; Mardon et al. 2023; Scholz and Smith 2019).

While most research emphasizes the positive outcomes of such relationships, such as persuasion, loyalty, and engagement (Leung et al. 2022), these same dynamics can also intensify harmful reactions. As group belonging is a powerful driver of adoption of toxic behaviors on social media (Zoizner and Levy 2025), when misinformation circulates parasocial closeness may amplify emotional defensiveness, leading to toxic or polarized exchanges. Influencers who deliberately court controversy or moral outrage to maintain visibility can further accelerate these dynamics (Coates et al. 2019; Barari 2023; Stewart et al. 2023). In extreme cases, parasocial identification has been linked to the diffusion of misogynistic or extremist narratives with offline consequences (Baele et al. 2024).

Because of parasocial relationships, influencers also occupy a different social and psychological status than regular social media users. Their audiences form cohesive, affective communities organized around admiration, trust, and shared values (Han and Balabanis 2024; Mardon et al. 2023). Within these communities, followers often internalize the influencer's perspectives, treating them as reliable authorities (Lou 2022). When misinformation originates from such figures, followers are not merely exposed to false content, but they are socially and emotionally invested in endorsing and defending it. This means that toxicity surrounding influencer-driven misinformation might frequently reflect belief alignment, as followers use hostile language to protect or promote the influencer's narrative against perceived outsiders or critics (Marwick and Boyd 2011).

By contrast, misinformation originating from regular users, who lack comparable credibility or parasocial influence, might elicit more fragmented and situational toxicity, often emerging from disagreement, ideological polarization, or anonymity rather than shared belief. In this sense, influencers do not simply attract more attention; they structure the emotional coherence of toxicity, transforming dispersed individual reactions into collective, belief-driven antagonism. Because parasocial ties transform passive audiences into emotionally engaged communities, misinformation introduced within these relationships is likely to generate toxicity that is more cohesive and belief-driven (Ekinci et al. 2025). Understanding whether such relational depth produces distinct forms of toxicity, compared with misinformation from regular users, is therefore central to the second research question.

RQ2. *What forms of toxicity emerge in response to brand-related misinformation originating from influencers compared with regular users?*

3.3 | Social Influence in Sustaining Toxic Echo Chambers

Social influence theory (Kelman 1958, 2006) provides an additional framework for understanding how influencer–audience dynamics evolve within online environments. Research has shown that influencer impact operates through three mechanisms of social influence: compliance (to gain approval), identification (to align self-concept), and internalization (when influencer values match personal beliefs) (Kelman 2006). Internalization, in particular, is driven by perceived credibility and expertise (McCormick 2016), allowing influencers to shape not only consumption decisions but also attitudes, values, and identity-related beliefs (Kapitan and Silvera 2016). These mechanisms explain why individuals adjust their attitudes and behaviors in response to others' expectations, relational bonds, or shared values, extending beyond persuasion to encompass how social belonging and identity maintenance shape reactions to communication.

In marketing, social influence has been widely applied to explain how opinion leaders and influencers drive attitude formation and behavioral conformity (Belanche et al. 2021). Influencers often elicit compliance through social approval cues such as likes or public endorsements, identification through aspirational self-presentation, and internalization when their messages align with followers' values and self-concept. While existing work acknowledges the emotional and identity-laden nature of brand interactions (Escalas and Bettman 2005; Monahan et al. 2023), little is known about how misinformation within these spaces is sustained and escalates to toxicity. Current studies largely examine individual-level effects such as trust erosion or reputational damage (Harrison-Walker and Jiang 2023; Lunardo et al. 2023), overlooking the social mechanisms through which toxic discourse spreads and reinforces itself.

These same mechanisms can also be at play at the intersection of misinformation and toxicity. When misinformation is disseminated by an admired or credible influencer, compliance may appear as mimicry or toxic defence to maintain group belonging (Mulcahy et al. 2024), identification may heighten

emotional sensitivity to perceived criticism, and internalization may entrench belief in misleading claims. As these mechanisms unfold in online environments, they may not only sustain persuasive influence but also enable the escalation and normalization of toxicity. Compliance, identification, and internalization can collectively transform isolated reactions into patterns of collective antagonism, particularly when followers mirror, defend, or embody an influencer's stance. Yet, little empirical attention has been given to how these mechanisms interact to perpetuate toxic discourse in brand-related misinformation contexts.

RQ3. *How do social influence mechanisms such as compliance, identification, and internalization drive the creation and amplification of toxicity surrounding brand-related misinformation shared by influencers?*

4 | Methods

To investigate how brand-related misinformation leads to toxicity in social media environments, we followed a multimethod empirical protocol combining top-down and bottom-up automated text analysis (Humphreys and Wang 2018) with qualitative thematic analysis. This approach was applied to a large-scale, cross-platform dataset derived from naturalistic, unmoderated social media environments where misinformation and hostile engagement are particularly prevalent (Cinelli et al. 2021; Schmidt et al. 2020).

We began by building a novel dataset focused on brand-related misinformation. Between 2020 and 2023, the fact-checking site *Snopes* published an estimated 5000–6000 verified misinformation items. From this large pool, we conducted a systematic keyword-based search using terms such as "consumer," "brand," "marketing," and "business" following keyword construction guidelines by Erdmann et al. (2022). The temporal scope (2020–2023) reflects both practical and theoretical considerations. Data collection was conducted in 2023, and the timeframe was defined to capture the 3 preceding years when misinformation activity and influencer engagement were at their peak. This period spans the COVID-19 pandemic and its aftermath with empirical evidence indicating that misinformation increased sharply during this time (Brennen et al. 2021), while influencer marketing activity and digital advertising investment also accelerated (Influencer Marketing Hub 2023). The inclusion of this timeframe also allowed the analysis to encompass the most active and relevant phase of misinformation circulation affecting brands.

The dataset primarily comprises content verified a US-based but internationally recognized fact-checking organization: Snopes. As such, the sample predominantly reflects Western digital environments, including English-language social media spaces where influencer marketing practices, platform affordances, and misinformation dynamics share strong commonalities. Prior studies show that influencer cultures in Western contexts operate through similar mechanisms of authenticity signaling and parasocial engagement (Zhu and Wang 2025), while misinformation circulates in comparable ways due to shared platform architectures and algorithmic amplification (Cinelli et al. 2021). Although some included brands have global reach, these shared communicative norms and digital infrastructures support the conceptual comparability of misinformation and toxicity dynamics across cases. A trained research assistant then manually screened these items to isolate those referencing identifiable brands in a consumer or marketing context. This process yielded 128 distinct misinformation cases referencing 47 brands across nine industries.

To verify the real-world circulation of these misinformation cases, we located the original misinformation post shared on social media, whether textual, visual, or video-based, and scraped the full postcontent, metadata, engagement statistics (e.g., likes, shares), and all associated audience comments. This allowed us to build a naturalistic, user-generated data corpus anchored in verified misinformation events.

The raw dataset initially included 128 posts, with over 105,453 associated audience comments. Following standard corpus preprocessing and data cleaning (Denny and Spirling 2018), we removed duplicate entries as well as non-English comments (Ooms 2023), replaced paralinguistic content, slang and word elongations (Rinker 2018). Also, to make sure that our comment corpus was not affected by potential trolls or social bots' activities, we conducted a thorough bot-detection analysis based on content, time, and account-based measures (Varol et al. 2017). The final dataset comprised 101 brand-related social media posts containing misinformation, originating from either SMIs or regular users, and 48,821 associated audience comments.

We employed a three-stage analytical protocol to address the research questions. First, to examine whether the source of misinformation (influencers vs. regular users) influences the intensity of toxic audience responses (RQ1), we conducted a comment-level logistic regression analysis. This model estimated the odds of a comment being classified as toxic while controlling for 15 post-, brand-, platform-, and comment-level variables. This was followed by post hoc analyses to determine conditions under which the source type shapes toxicity. Second, to explore how the forms and expressions of toxicity differ depending on post source (RQ2), we combined topic modeling with regression analysis. This stage identified five distinct discursive categories of toxic comments, capturing variation in emotional tone, target, and intent across influencer- and user-generated misinformation.

Finally, to understand how influencers contribute to the creation and amplification of toxicity through social influence mechanisms (RQ3), we conducted an in-depth thematic analysis of 1800 representative toxic comments and a subsample of SMI-generated posts ($n = 45$). Using a hybrid coding approach that integrated theory-informed and inductive procedures (Braun and Clarke 2006; Fereday and Muir-Cochrane 2006), we traced how misinformation narratives are discursively constructed, endorsed, and sustained within influencer-led interactions. The overall analytical process is summarized in Figure 1.

5 | Top-Down Automated Text Analysis

5.1 | Procedure

To identify toxic responses within the data set, we used Google's Perspective API (Perspective AI 2024) to analyse audience comments associated with each brand-related misinformation post. Given the dataset's size, the API offered an efficient and scalable

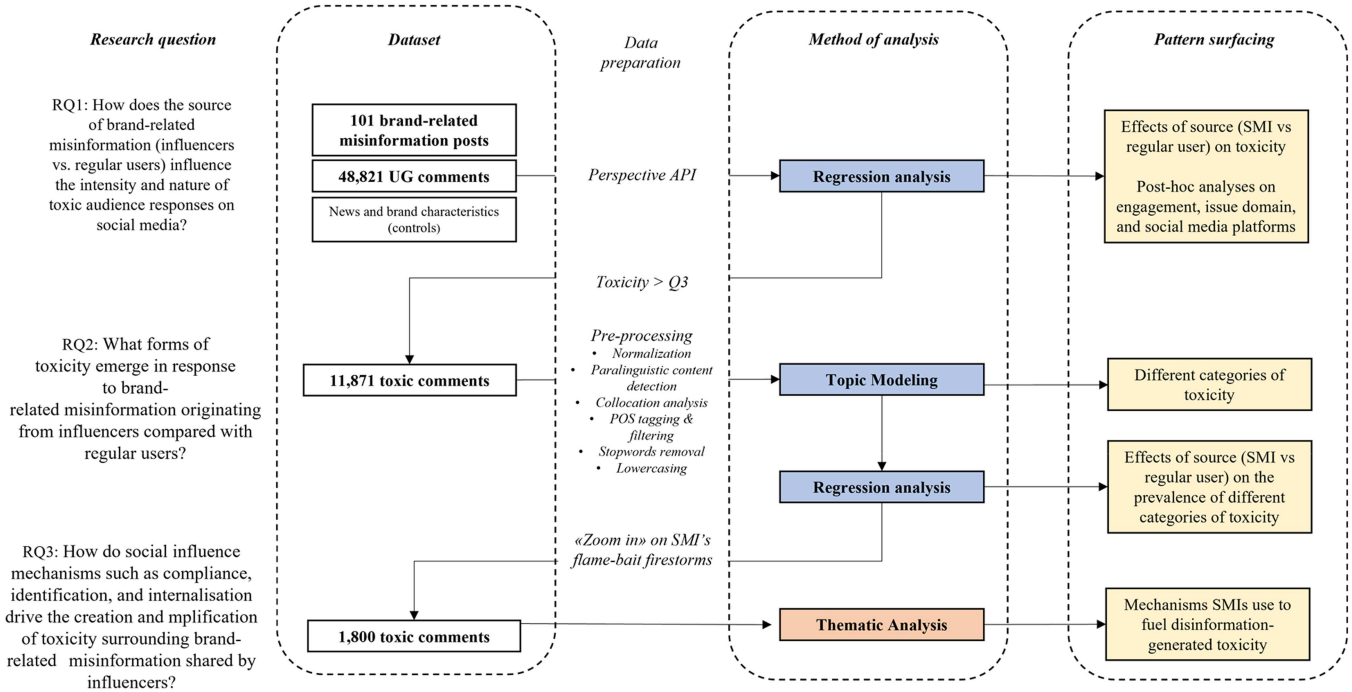

FIGURE 1 | Analytical process.

TABLE 2 | Summary of validity checks.

Type of validity	Description
Construct	Following Humphreys and Wang (2018), for the toxicity score (RQ1), we used a top-down method employing the state-of-the-art deep learning toxicity detection algorithm Perspective API. For the categories of toxicity (RQ2), we employed a bottom-up approach, combining topic modeling and thematic analysis.
Concurrent	The automated toxicity score shows substantial concurrence with human ratings (average pairwise percentage agreement = 0.83).
Convergent	The toxicity measure shows a strong, positive correlation with Detoxify ($r = 0.75$, $p < 0.001$; Hanu and Unitary team 2020), and a moderate, positive correlation with Hurtlex ($r = 0.34$, $p < 0.001$; Bassignana et al. 2018).
Discriminant	The toxicity measure does not show strong, significant correlations with other constructs included in the model (see Table SA2, Column 1).
Causal	We included 15 control variables to account for alternative explanations (see Table 3).
Predictive	The theoretically derived relationship between toxicity and source type was replicated across multiple robustness checks.
Face	The automated classification aligns with intuitive human judgments. For instance, a toxic comment such as "F**k you @brand once a racist company always one!" contrasts clearly with a civil comment such as "So sad that people feel they have a right to say how another person should love and feel. Thank God things are starting to change. Love is love < 3."

method for detecting toxicity in user-generated text. The tool defines toxicity as "a rude, disrespectful, or unreasonable comment that is likely to make someone leave a discussion" (Perspective AI 2024). This operationalisation aligns closely with this study's focus on hostile and exclusionary discourse, capturing the broader contours of user toxicity in social media environments.

The Perspective model has demonstrated high reliability, with accuracy rates above 90% (Lin et al. 2024), and has been validated across domains including news, gaming, and marketing (e.g., Avalle et al. 2024; Nepomuceno et al. 2023). We ran the Perspective "AnalyzeComment" API directly on our corpus of audience comments, which assigns each text a toxicity score ranging from 0 (*nontoxic*) to 1 (*highly toxic*). For example, a score of 0.6 indicates that six out of ten human coders would classify the comment as toxic. Following prior research (Avalle et al. 2024), we adopted a 0.6 threshold to generate a binary dependent variable coded as toxic (1) if the score exceeded this threshold and nontoxic (0) otherwise. Several validation checks confirmed the robustness of this text-based toxicity measure (see Table 2 and Appendix S1A for more details).

To address RQ1, we examined whether the source of misinformation, specifically, whether a post was authored by an influencer or a regular user, affects the odds of eliciting toxic audience responses. Our main explanatory variable, source, was constructed using a hybrid classification approach that combined profile-level and content-based analysis. First, we assessed account-level indicators such as bio descriptions, follower counts, engagement metrics, and thematic consistency to determine whether a user projected an influencer identity. We then analyzed the linguistic and visual characteristics of users' posts, focusing on tone, imagery, and self-presentation strategies. Intercoder reliability was high across multiple rounds ($\kappa = 0.89$; $\kappa = 0.97$), following established influencer classification frameworks (Bonini et al. 2016; Caliandro and Gandini 2016).

We also controlled for 15 post-, brand-, platform-, and comment-level variables that may influence toxicity (see Table 3). Given the binary outcome variable (toxic vs. nontoxic), we employed logistic regression models to estimate the probability of a toxic response as a function of source type and control variables. This model enabled a systematic test of whether influencer-generated misinformation is more likely to trigger toxic engagement than misinformation shared by regular users (Miranda et al. 2022; Schmidt et al. 2020).

5.2 | Results

Table 4 presents the results of our logistic regression analyses. For clarity, we report exponentiated coefficients (odds ratios). Model 1 includes only control variables and establishes the baseline likelihood of toxic comments across the dataset. Model 2 adds our key predictor, post source, to test whether toxicity levels differ depending on whether the misinformation was shared by a SMI or a regular user.

Results show that, holding all the other variables constant, posts authored by regular users are significantly less likely to generate toxic comments than those posted by SMIs. Specifically, the odds of toxicity below user-generated content is 44% lower than those of SMI (odds ratio = 0.56, $p < 0.001$). Marginal effects indicate that the predicted probability of toxicity is approximately 3.8% for SMI posts and 2.2% for user posts, a difference of 1.6 percentage points, reflecting a substantively meaningful reduction (Schmidt et al. 2020). In other words, misinformation coming from SMIs is more likely to spark toxicity in the comment sections.

The negative effect of source on toxicity is robust and extends beyond correlational evidence. Across a series of modeling and measurement robustness checks (Table 5), the effect remains stable in magnitude and significance after accounting for selection bias, endogeneity, and alternative measures of both the dependent and independent variables. This consistency across specifications supports that the source effect reflects a substantive, not spurious, relationship. Robustness analyses confirmed that results were stable across variations in corpus size, model specification, and controls for selection bias and endogeneity. More details about the robustness analyses can be found in Appendix S1C–S1I.

5.3 | Post Hoc Analyses

To better understand how and when source type shapes toxicity, we conducted a set of follow-up analyses (illustrated in Figure 2). We first considered engagement as a boundary condition. Engagement is central to influencer economies (Leung et al. 2022), where visibility depends on audience reactions. It can also reward emotionally charged or morally provocative content (Avalle et al. 2024). Because influencers actively manage these content–reaction loops, engagement may shape toxicity differently for them than for ordinary users.

The interaction between source type (SMI vs. user) and engagement was significant (OR = 0.81, 95% CI [0.76, 0.86], $p < 0.001$; see Figure 2). For SMIs, engagement increased with toxicity (OR = 1.07, 95% CI [1.05, 1.09], $p < 0.001$), while for users it declined (OR = 0.87, 95% CI [0.85, 0.89], $p < 0.001$). This crossover indicates a toxicity–engagement spiral unique to influencers: their more toxic posts attract stronger reactions, reinforcing incentives to post such content. Regular users show the reverse pattern, where engagement aligns with more civil exchanges.

We next examined whether this pattern varies by issue domain. Although all posts contained brand-related misinformation, they addressed different topics—commercial, sociopolitical, or health and safety—which differ in salience and psychological distance (Trope and Liberman 2010). The interaction between source type and issue category was significant. Pairwise contrasts (FDR-adjusted) revealed that influencers triggered more toxicity than users only in sociopolitical discussions (OR = 1.87, 95% CI [1.53, 2.28], $p < 0.001$). Differences were nonsignificant for commercial ($p = 0.085$) and health and safety ($p = 0.85$) content. Influencers thus appear to amplify toxicity particularly when misinformation concerns socially charged issues.

Finally, we tested whether the platform environment moderates this relationship. Platforms vary in affordances and norms that shape communicative behavior (Literat and Kligler-Vilenchik 2021). We grouped them by pseudonymity: low (Meta), medium (YouTube, TikTok), and high (Reddit, X). A significant interaction emerged. Influencers were more toxic than users on all platform types (Low: OR = 2.10, 95% CI [1.75, 2.52]; Medium: OR = 1.97, 95% CI [1.36, 2.86]; High: OR = 4.81, 95% CI [2.56, 9.02]; all $ps < 0.001$). Among influencers, toxicity peaked on low-pseudonymity platforms (OR = 1.36, 95% CI [1.11, 1.68], $p = 0.003$). User differences across platforms were smaller, with only the low–medium contrast significant (OR = 1.62, 95% CI [1.05, 2.51], $p < 0.05$). These results suggest that influencer toxicity is not confined to anonymous or unregulated environments but may be most visible where identity and reputation are salient.

6 | Bottom-Up Automated Text Analysis

6.1 | Topic Modeling

We began this stage by identifying and characterizing key forms of toxicity through a dimensionalisation process (Miranda et al. 2022). Focusing on the most extreme cases, we analyzed the top quartile of comments by toxicity score ($n = 11{,}871$) to capture the diverse expressions toxicity assumes in online exchanges.

TABLE 3 | Variable definition and descriptive statistics.

Construct	Definition	Measurement strategy	Mean (standard)	Median [minimum, maximum]
Dependent variable				
Toxicity	"a rude, disrespectful, and unreasonable comment that is likely to make someone leave a discussion" (Perspective AI 2024)	ATA- A comment is labeled as toxic if toxicity score > 0.6 (Avalle et al. 2024).	0.11 (0.40)	0 [0, 1.00]
Independent variable				
Source	Source of the post	Analysis of the profiles of and of the self-presentation strategies enacted by the source (Caliandro and Gandini 2016)	0.29 (0.45)	0 [0, 1.00]
Controls				
News truthfulness	Veracity evaluation of news after their first circulation (Visentin et al. 2019)	Snopes rating	0.23 (0.42)	0 [0, 1.00]
Issue	Nature of the issue that the post leveraged upon in its claim (source)	Manual content analysis of the posts	3.10 (1.18)	4.00 [1.00, 4.00]
Industry	Sector any brand mentioned in a given post competes in (Shahbaznezhad et al. 2021)	—	3.24 (2.16)	3.00 [1.00, 9.00]
Political affiliation	Political affiliation of any brand mentioned in a given post	Preference partisanship brand score (Schoenmueller et al. 2022). Membership is computed via median split.	1.86 (0.71)	2.00 [1.00, 3.00]
Year	Year when a specific news was posted (Shahbaznezhad et al. 2021)	—	2.57 (1.05)	3.00 [1.00, 4.00]
SM platform	Social media platform where a specific news was posted (Shahbaznezhad et al. 2021)	—	2.71 (1.89)	2.00 [1.00, 6.00]
Comments length	Comments' length measured by word count	ATA (word count)	21.9 (24.1)	16.0 [1.00, 1910]
Engagement effect	Effect of low-involvement engagement behavior on high involvement engagement behavior and vice versa, for the same post (Shahbaznezhad et al. 2021)	Sum of the engagement metrics (e.g., likes, comments, shares) accumulated by each post (log-transformed)	11.2 (1.80)	11.3 [1.39–15.0]
Emotionality- anger	Emotional orientation of the posts (Brady et al. 2023)	Intensity of basic emotions expressed in each post's text, measured as a probability score via NADE emotional detector (Hotz-Behofsits et al. 2025)	0.11 (0.07)	0.09 [0, 0.48]
Fear	Fear		0.11 (0.08)	0.09 [0, 0.47]
Anticipation	Anticipation		0.24 (0.09)	0.27 [0, 0.43]
Disgust	Disgust		0.17 (0.10)	0.17 [0, 0.47]
Sadness	Sadness		0.09 (0.06)	0.08 [0, 0.39]
Joy	Joy		0.19 (0.09)	0.19 [0, 0.55]

Note: Industry (Percentage of the data set): Automotive (16.7), Big Tech (32.2), Food and Beverages (27.8), Financial (0.1), Luxury and fashion (0.6), News (0.1), Retail (11.0), Tech (5.8), Telcom (0.8); Issue: Commercial (14.3), Health and safety (22.5), Other (2.4), Sociopolitical (60.9); Posting Year: 2020 (20.1), 2021 (25.9), 2022 (30.8), 2023 (23.2); Social media platforms: Facebook (43.6), Instagram (18.7), Reddit (0.4), TikTok (6.5), Twitter (22.3), YouTube (8.5); Political affiliation: Democrat (33.6), Liberal (47.1), Unknown (19.3); Source: SMI (0.71); News truthfulness: True (22.5); The basic emotion "surprise" was removed from the analyses due to its low scores. $N = 48{,}821$.

TABLE 4 | Logistic regressions: Effect of source on toxicity.

DV: Toxicity	(1)		(2)	
Predictors	**Odds ratios**	**Standard error**	**Odds ratios**	**Standard error**
SOURCE [user]			0.56***	0.06
Controls				
News truthfulness [TRUE]	0.85	0.11	0.72*	0.10
Issue [Health and safety]	1.26*	0.12	1.15	0.12
Issue [Other]	0.13***	0.03	0.09***	0.02
Issue [Sociopolitical]	0.55***	0.05	0.52***	0.05
Posting year [2021]	0.28***	0.02	0.23***	0.02
Posting year [2022]	0.76***	0.06	0.83*	0.07
Posting year [2023]	0.33***	0.03	0.30***	0.03
Engagement effect	1.00	0.02	1.00	0.02
Anger	1.21	0.54	4.80**	2.49
Fear	0.84	0.39	0.26*	0.14
Trust	0.48**	0.11	0.32***	0.07
Anticipation	0.02***	0.01	0.02***	0.01
Disgust	182.86***	72.88	137.59***	56.34
Sadness	0.02***	0.01	0.04***	0.02
Joy	0.48*	0.15	0.31***	0.10
Political affiliation [Liberal]	1.08	0.10	1.16	0.11
Political affiliation [Unknown]	0.48***	0.06	0.64***	0.09
Industry [Big Tech]	1.04	0.11	1.44**	0.18
Industry [Food and Beverages]	0.65***	0.06	0.73**	0.07
Industry [Financial]	0.00	0.01	0.00	0.01
Industry [Luxury and fashion]	1.13	0.20	1.70**	0.33
Industry [News]	1.62	0.93	1.20	0.70
Industry [Retail]	1.02	0.09	1.12	0.11
Industry [Tech]	0.66**	0.10	0.93	0.15
Industry [Telcom]	1.85*	0.48	1.82*	0.48
SM platform [Instagram]	2.63***	0.22	1.88***	0.19
SM platform [Reddit]	9.20***	2.78	12.99***	4.02
SM platform [TikTok]	1.10	0.15	0.85	0.12
SM platform [Twitter]	1.08	0.07	0.82*	0.07
SM platform [YouTube]	0.34***	0.05	0.33***	0.05
Comment thread's length	1.00***	0.00	1.00***	0.00
(Intercept)	0.55*	0.13	0.78	0.18
Observations	48,821		48,821	
LR χ^2 (*df*)			34.30 (1)***	
R^2 Tjur	0.038		0.039	

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

To uncover recurring discursive patterns within these highly toxic comments, we applied the Biterm Topic Model (BTM), a version of Latent Dirichlet Allocation optimized for short texts (Yan et al. 2013). The dataset was preprocessed (tokenisation, collocation analysis, part-of-speech tagging, stop-word removal, stemming), and model parameters were iteratively tuned to maximize coherence. The final model produced 48 topics ($\alpha = 1.04$, $\beta = 0.01$). To ensure conceptual precision, all topics underwent human verification: each was independently reviewed and labeled by the research team based on top keywords and representative comments. Following established validation procedures (Aranda et al. 2021; Greve et al. 2022), topics were assessed for semantic coherence and distinctiveness, with thematically weak outputs excluded. This process yielded 41 coherent topics, 36 of which reflected 5 distinct toxic discourses.

TABLE 5 | Summary of the modeling and measurement robustness checks.

Category	Robustness check	Method and outcome
Modeling	1. Size vs. spurious statistical significance	Coefficient/p value/sample-size (CPS) chart (Lin et al. 2024). CPS chart for the primary explanatory variable shows little sensitivity of coefficient and p value as sample size increases (Appendix S1B).
	2. Selection bias and correlational evidence	Propensity score matching. Odds of a comment being toxic are ~39% lower for regular user posts than influencers (OR = 0.61, 95% CI [0.54, 0.69]).
	3. Endogeneity of source	Gaussian copula. Not significant copula term (GC: 0.16, SE: 0.08, $p > 0.05$)
	4. Alternative DV	Using "insult" and "threat" as DVs; OR = 0.58, $p < 0.001$—Model A3; OR = 0.17, $p < 0.05$—Model A4.
Measurement	1. Alternative IV	Replacing source with SMI taxonomy (Campbell and Farrell 2020): nano and micro SMIs show higher toxicity than macro SMIs; OR micro = 1.27, $p < 0.05$; OR macro = 0.70, $p < 0.05$; OR mega = 1.93, $p < 0.001$—Model A6.
	2. Sensitivity analysis for toxicity	Using perspective API continuous score: OLS regression: β source = −0.27, $p < 0.001$—Model A5.

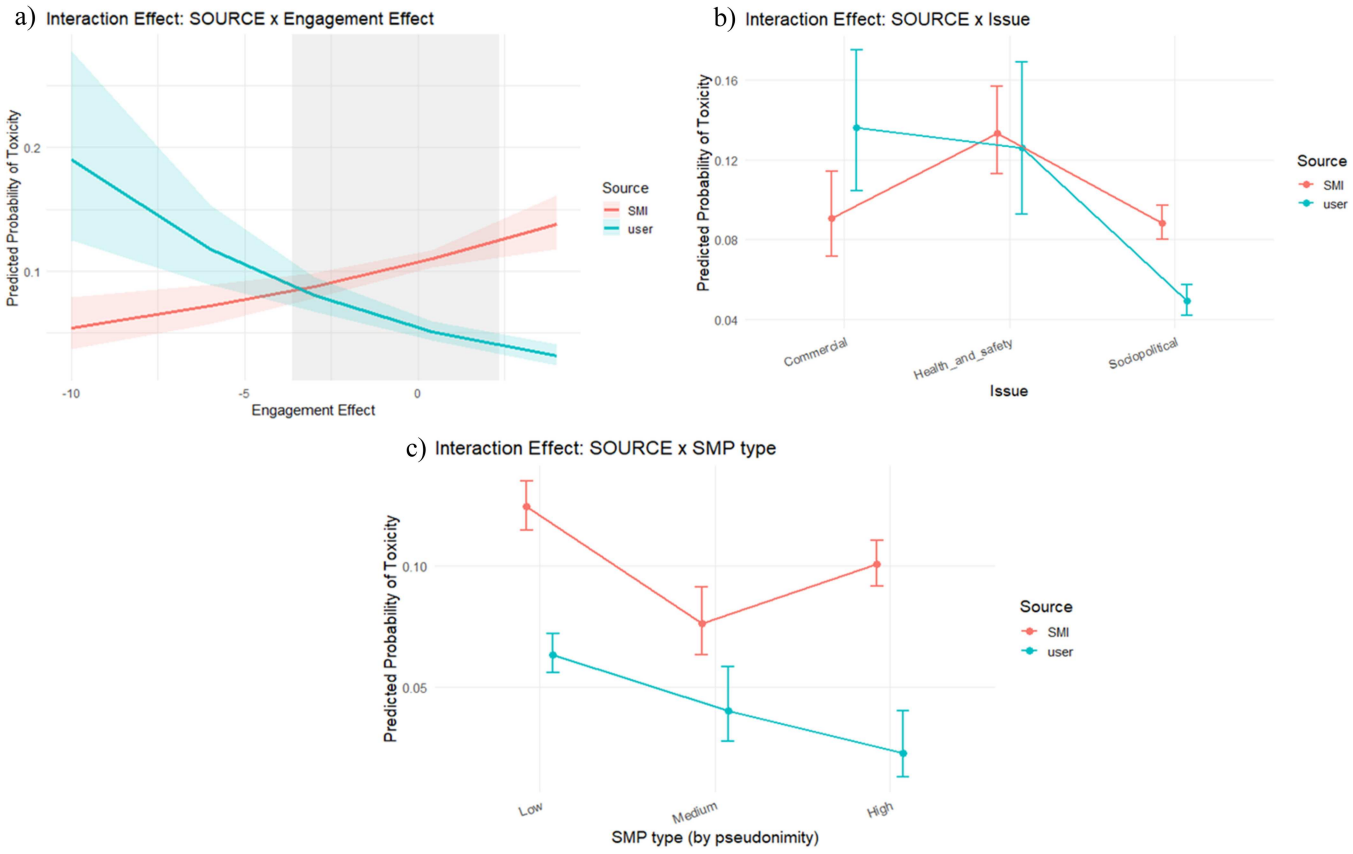

FIGURE 2 | Conditional effects of source type on toxicity across engagement (a), issue domain (b), and platform (c). (a) Engagement effect was centralized before modeling; shaded area indicates where 95% of the observations lie. Panel b) Issue category "other" was removed from this analysis ($N = 47{,}663$).

From these 36 validated topics, we developed a higher-order taxonomy capturing the main modes of interactive toxicity in audience responses to brand-related misinformation. Building on prior research, topics were categorized by discursive focus (brand- vs. consumer-directed), behavioral intent (venting, retaliation, provocation, or deliberate harm), and relational target (brands, other consumers, or wider audiences). This synthesis incorporated established forms of online toxicity, including malicious negative word-of-mouth (Hornik et al. 2019; Liao et al. 2024), firestorms (Herhausen et al. 2019), anti-brand activism (Romani et al. 2015), trolling and flaming (Dineva and Breitsohl 2022), harassment (Vogels 2021), and consumer-to-consumer conflict or brand bullying (Dineva and Daunt 2023). These classifications formed the analytical framework for coding and interpreting toxicity in the dataset (see Appendix S1J for the bottom-up analysis summary table).

6.2 | Regression Analyses

To examine how toxicity categories vary depending on the source of social media content, we conducted a category prediction task (Miranda et al. 2022). Specifically, we ran 5 separate ordinary least squares (OLS) regressions with robust standard errors and FDR adjustments, one for each category of toxicity identified through our earlier modeling. In each model, the log-transformed prevalence of the category served as the dependent variable, with content source (SMI vs. regular user) as the key independent variable, alongside relevant controls.

This approach allowed us to systematically estimate how the presence and intensity of specific toxic categories differed by source type. By modeling category prevalence as a function of source, we were able to assess the extent to which influencer-generated content is associated with distinct patterns of toxic audience responses, compared to those emerging from regular users.

6.3 | Results

We identified five distinct categories of toxic interactive behaviors enacted by users in response to brand-related misinformation on social media: "anti-brand reactions," "C2C conflicts," "flame-bait firestorms," "toxic debunking," and "trolling and flaming."

The first category, termed "anti-brand reactions," encompasses toxic comments that specifically target a brand, its activities, or its supporters. These reactions vary in severity, from mild cynicism about a brand's consumer choices (e.g., topic 35) or schadenfreude towards failed brand innovations (e.g., topic 2) to more intense forms of toxicity such as calls for boycotts (e.g., topic 38) and outright brand-bullying (e.g., topic 41). The primary targets of these comments are either the brand in question or other consumers, particularly those defending or supporting the brand with an emphasis on creating attacks directed at a brand (Dessart et al. 2020).

The second category, labeled "C2C conflicts," refers to violent exchanges between consumers who hold opposing views on contentious consumption-related issues, such as minority representation (e.g., topic 3) or food safety (e.g., topic 11). Unlike anti-brand reactions, which are often directed at the brand itself or other users, C2C conflicts involve direct confrontations between users. These interactions are characterized by a dialogical yet antagonistic exchange of accusations, leading to a vindictive form of toxicity (Garimella et al. 2017). This toxicity category is concerned with interpersonal disputes and polarization, rather than the involved brand or the content of the misinformation news (Dineva and Daunt 2023; Luedicke et al. 2010).

The third category, "flame-bait firestorms," consists of highly emotional and unified reactions against the brand or news itself rather than against individual users. This category is marked by intense backlash or firestorms aimed directly at the news' focal point, such as the brand implicated in the controversy (Herhausen et al. 2019; Scholz and Smith 2019). Unlike exchanges in C2C conflicts, which are characterized by multi-directional and polarized interactions involving a wide range of perspectives, flame-bait firestorms create a very different kind of discourse. In flame-bait firestorms, the discussions tend to become highly homogeneous and monologic. This means that the conversations often reinforce a singular viewpoint and are resistant to incorporating or even acknowledging opposing perspectives. Hence, this toxicity category entails discourse that creates a reinforcing loop where only similar opinions are circulated and amplified, making it difficult for dissenting voices or alternative viewpoints to penetrate the discussion (Cinelli et al. 2021).

The fourth category, "toxic debunking," aggregates comments that aim to discredit or ridicule problematic news content through verbal aggression. This category features two main debunking strategies: *low-involvement debunking*, characterized by direct insults and personal attacks without substantive evidence (e.g., topic 21), and *high-involvement debunking*, where users provide logical counterarguments supported by personal experience or external sources (e.g., topic 20). Both forms of debunking target the source of misinformation (i.e., SMIs or regular users), whether through personal attacks or reasoned refutations. This category of toxicity differs significantly from C2C conflicts. In this context, the focus is on debunking misinformation and verifying its authenticity. In contrast, C2C conflicts are primarily concerned with sustaining polarization and do not prioritize reaching a consensus on the alleged validity of the information.

The fifth category, "trolling and flaming," includes a range of toxic behaviors designed to provoke reactions for the troll's amusement or emotional release (Cho and Kwon 2015; Golf-Papez and Veer 2022). This category encompasses both subtle provocations aimed at mocking users or brands (e.g., topic 43) and more overtly offensive insults conveyed through aggressive language (e.g., topic 25). Trolling and flaming are often marked by their superficial nature and lack of a specific target, revealing a broader intent to provoke reactions rather than engage in content verification, as seen in toxic debunking. Unlike C2C conflicts, where the goal is to sustain existing polarization, trolling and flaming aim to create polarization from the outset. Furthermore, instead of adhering to a consistent "echo-chamber" viewpoint, as in flame-bait firestorms, trolling and flaming adopt whichever stance is most likely to generate toxicity.

Our category-level regression analyses (Figure 3) reveal a compelling pattern: comments on regular users' accounts display a greater diversity of toxic behaviors and targets compared to those on SMIs' posts, visually represented i by the larger number of bubbles, each with distinct colors and targets (y-axis). Notably, regular users' posts tend to provoke more out-group toxic behaviors, particularly in anti-brand controversies (Std coefficient: 0.94, $p < 0.001$), where negative word-of-mouth, rivalries, and boycott intentions dominate the discussions. Moreover, regular user comments show a higher prevalence of toxic debunking instances (standard coefficient: 0.61, $p < 0.001$), suggesting more dialogic forms of toxicity (Scheibenzuber et al. 2023). In contrast, the comment sections of SMIs are more homogeneous, largely dominated by toxic reactions that fit within the "flame-bait firestorms" category (standard coefficient: −1.13, $p < 0.001$), where collective consumer animosity is directed at a target (i.e., the brand victim of misinformation) unrelated to the influencer themselves. These findings suggest a marked difference in the types of toxicity provoked by regular

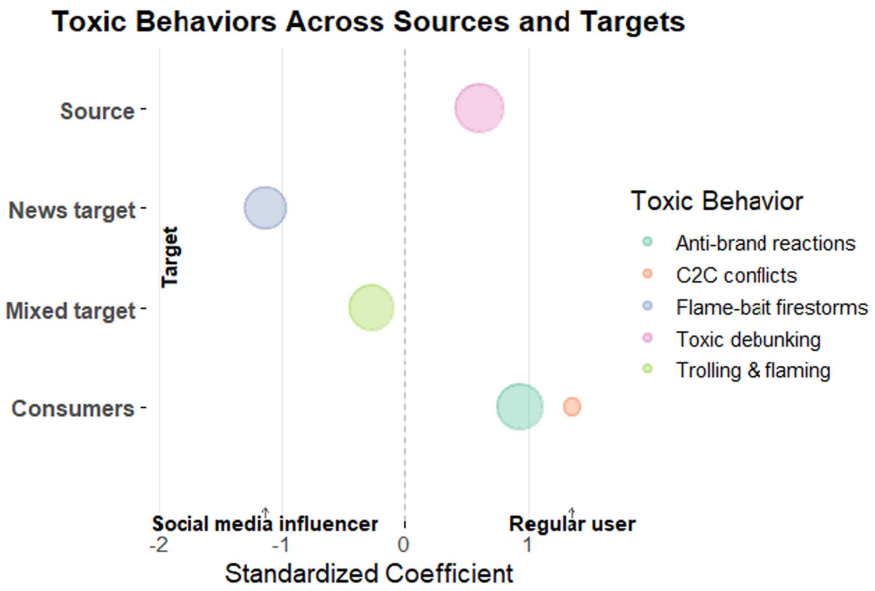

FIGURE 3 | Consumer toxic categories by source and target. The *x*-axis displays the standardized coefficients from category-level regressions of *source type* on log-transformed category prevalence, controlling for political affiliation, issue domain, social media platform, and posting year (see Appendix S1K). Categories that appear significantly more often in comment sections under influencers' posts are positioned on the left, whereas those more prevalent under regular users' posts appear on the right. The *y*-axis indicates each topic's primary target. Bubble size reflects category prevalence, and colors represent the specific toxic discursive behavior each category embodies.

users versus SMIs, with regular users facing a broader spectrum of harmful online behavior.

Interpreted through parasocial interaction theory, this pattern suggests that followers form one-sided emotional bonds with influencers (Horton and Richard Wohl 1956), which shape how they react to misinformation content. Because these parasocial relationships foster perceived intimacy, authenticity, and loyalty (Labrecque 2014), audiences tend to protect the influencer rather than challenge them. As a result, toxic engagement is redirected toward a safer external target—the brand implicated in the misinformation—rather than toward the influencer who initiated or amplified it. In contrast, when misinformation originates from regular users, no parasocial buffer exists to shield them from scrutiny or backlash, making them more vulnerable to a broader range of harmful interactions, including direct attacks and credibility questioning (Chung and Cho 2017).

7 | Mechanisms of Influencer-Driven Toxicity

To investigate how brand-related misinformation gives rise to toxic audience behaviors and the role of SMIs in shaping these dynamics, we conducted a thematic analysis (Braun and Clarke 2006) of a purposive subsample of user-generated content. In doing so, we adopted a hybrid coding strategy that combined theory-driven and inductive approaches (Fereday and Muir-Cochrane 2006). We first applied a predefined coding template based on existing research on SMI engagement strategies (see Section 2.2). These included: "credibility signalling" (Kapitan and Silvera 2016; McCormick 2016), "parasocial bonding" (Thomas et al. 2024; Reinikainen et al. 2020), "strategic controversy amplification" (Barari 2023; Stewart et al. 2023), "algorithmic manipulation" such as comment pinning and thread boosting (Feng and Kim 2024), and "populist narrative framing" using "us vs. them" discourse (Holt and Cameron 2010; Fong et al. 2021). Irrelevant codes were excluded based on dataset fit.

Next, we conducted inductive coding to identify context-specific patterns that existing frameworks did not capture. This allowed us to surface SMI discursive strategies unique to brand-related misinformation including emergent behaviors such as "pinning to reinforce narrative salience," "rehashing past controversies," "strategic silence," "emotional self-disclosure." and "follower commending." Throughout, we iteratively refined both theory-based and emergent codes, clustering them into broader thematic categories (Fereday and Muir-Cochrane 2006). Finally, we synthesized the coding structure into a cohesive thematic framework. This included an interpretive layer focused on identifying the underlying mechanisms SMIs used to provoke and sustain engagement. In analyzing SMI posts and comment sections, we paid close attention to how influencers presented themselves, framed misinformation, and interacted with their audiences (Cocker and Cronin 2017).

Throughout the process, coding was conducted iteratively by multiple researchers. Agreement was calculated across three rounds ($\kappa = 0.87$; $\kappa = 0.91$; $\kappa = 0.96$), with discrepancies resolved through discussion and refinement of the coding approach (Milne and Adler 1999). The results provided a framework for identifying the influencer strategies that shape patterns of toxic behaviors among social media users.

7.1 | Results

Our qualitative analysis uncovered the mechanisms that SMIs adopt to cultivate toxic echo chambers in response to brand-related misinformation on social media through two distinct mechanisms, as illustrated in Figure 4.

The first mechanism, *brand-related misinformation legitimation*, captures how influencers validate and protect misinformation

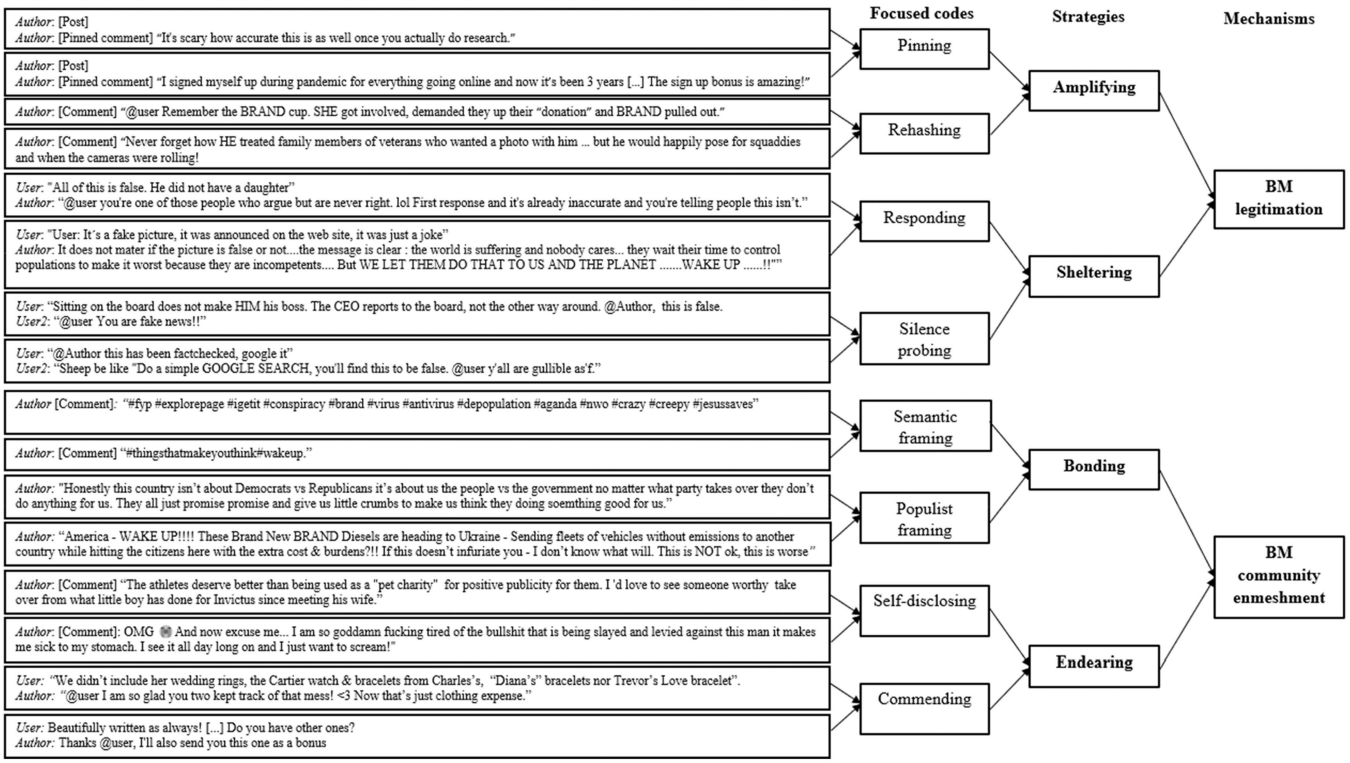

FIGURE 4 | Mechanisms through which social media influencers foster toxic echo chambers around brand-related misinformation.

central to their campaigns (Di Domenico et al. 2022). This mechanism reflects compliance-based influence (Cialdini and Goldstein 2004; Kelman 2006): followers are prompted to publicly endorse misinformation to preserve group belonging and avoid backlash. Two main strategies underpin this process.

Amplifying increases misinformation's reach and perceived legitimacy by exploiting engagement affordances. Influencers strategically interact in comment threads to highlight supportive remarks and create an illusion of consensus (Feng and Kim 2024). For instance, under a false claim that a multinational's skincare products "secretly contain banned chemicals," one influencer replied, "Exactly - people have no idea what's really in this stuff," which drew hundreds of affirming responses and extended the conversation over several days. Influencers also rehash misinformation by adding details or reviving old controversies, for example, revisiting a debunked rumor that a celebrity boycott had forced a brand recall, to sustain outrage and maintain visibility.

Sheltering focuses on defending misinformation from scrutiny. When challenged, influencers reframe criticism as hostility, often mobilizing followers to respond on their behalf (Timothy Coombs and Jean Holladay 2014; Stieglitz et al. 2019). In one thread, after a user questioned a post's accuracy, the influencer replied, "@user you clearly work for them-stop gaslighting people," prompting numerous supporters to attack the critic. Others employ "strategic silence," allowing followers to defend the influencer without direct engagement. Across cases, these behaviors transform informational credibility into social power, reinforcing the influencer's authority and discouraging dissent.

The second mechanism, *community enmeshment*, reflects how influencers cultivate emotionally bonded, homogeneous communities that act as self-reinforcing echo chambers (Rao and Greve 2024). This mechanism operates through identification and internalization processes (Kelman 1958). *Bonding* aligns followers with the influencer's worldview through in-group rhetoric and moral framing. Influencers regularly use populist and conspiratorial language (e.g., "#wakeup," "they're hiding the truth") to reinforce shared identity and collective purpose (Fong et al. 2021; Marwick and Boyd 2011). When one influencer wrote, "It's not about left or right pp it's about us versus the corporations," followers echoed the sentiment, generating a long thread of agreement and anti-brand hostility.

Endearing strengthens parasocial attachment through emotional self-disclosure and personalized recognition (Labrecque 2014). Influencers share vulnerable or affective posts, such as expressing exhaustion over being "attacked for telling the truth," to evoke empathy and solidarity. They also publicly praise supportive users ("@user You've been here since day one - thank you for fighting with me"), reinforcing intimacy and loyalty. These displays foster perceived closeness and mutual defence, embedding misinformation within a shared emotional and moral narrative.

Combined, brand-related misinformation legitimation and community enmeshment explain how influencers convert misinformation into collective toxicity. The first mechanism reinforces misinformation through visibility and defence, leveraging compliance and social proof; the second embeds it through emotional resonance and identity alignment, driving internalization of misinformation as group truth. The two strategies illustrate how relational authority and networked influence transform misinformation from isolated content into a socially sustained, toxic dynamic that amplifies hostility and discourages correction.

8 | Discussion

This paper addresses the largely overlooked role of SMIs in the spread of misinformation and online toxicity. Our analysis shows that brand-related misinformation shared by SMIs, compared with regular users, provokes significantly more toxic audience reactions. This pattern aligns with source credibility theory, as influencers' perceived authority and trustworthiness heighten responsiveness and lower scepticism, increasing the likelihood of hostile engagement. Post hoc analyses reveal that influencers not only generate more toxicity but amplify it under the same conditions that enhance their visibility and influence. Toxicity escalates with engagement, producing a self-reinforcing toxicity–engagement spiral; it peaks when influencers discuss socio-political issues, where public stakes are higher; and it is most pronounced on low-pseudonymity, identity-based platforms that reward visibility and reputation.

Viewed through parasocial interaction theory, these effects reflect qualitative differences in how toxicity unfolds. Followers' one-sided emotional bonds with influencers (Horton and Richard Wohl 1956; Labrecque 2014) foster loyalty and perceived intimacy, prompting defence rather than critique, even when misinformation is apparent. This loyalty redirects hostility toward external targets, typically the brand implicated in misinformation, generating homogeneous flame-bait firestorms aligned with the influencer's narrative. In contrast, misinformation from regular users lacks such relational protection, exposing them to direct attacks and credibility challenges (Chung and Cho 2017). Toxicity in influencer-initiated conversations is therefore not only greater in magnitude but also shaped by relational dynamics that shield influencers and displace aggression. Finally, our thematic analysis shows that SMIs employ discursive and technical tactics that activate compliance, identification, and internalization, core social influence processes, sustaining toxic echo chambers around misinformation. Influencers thus become catalysts of online toxicity, shaping both its intensity and its direction.

8.1 | Theoretical Contributions

This research advances understanding at the intersection of misinformation, influencer marketing, and online incivility in three ways. First, this research advances the influencer marketing literature by demonstrating that misinformation disseminated by SMIs produces more, and qualitatively different, toxicity than identical content from regular users. Rather than attributing harm solely to message content, the findings theorize toxicity as an outcome of source-based amplification. Drawing from source credibility theory (Hovland and Weiss 1951), we find that SMI's perceived expertise, visibility, and platform positioning activate credibility cues that legitimize misinformation (Di Domenico et al. 2022) and lower normative barriers to incivility. In this respect, the study extends prior discussion of problematic influencer conduct (Coates et al. 2019; Karagür et al. 2022) by showing how credibility itself becomes an infrastructural mechanism through which toxicity is normalized. Practices such as content pinning or selective interaction function not only as engagement-maximizing tactics, but as symbolic endorsements that stabilize misinformation narratives and atmospheres (Bahar and Hasan 2024; Mangiò and Di Domenico 2022). The identified boundary conditions—issue salience, engagement incentives, and low pseudonymity—further support this original stance where credibility shifts from a persuasive resource into a catalyst for harm.

Second, the study reconceptualises online toxicity as a relational and performative outcome of misinformation, expanding dominant marketing perspectives that focus primarily on attitudinal or trust-based effects (Di Domenico and Ding 2023). Drawing on parasocial interaction and social influence perspectives, the findings show that influencer-led misinformation does not simply intensify incivility but reorganizes its expression. The proposed typology distinguishes between established forms of toxicity—such as consumer conflict, trolling, flaming, and antibrand activism—and emergent patterns that are specific to influencer-centered contexts. While the former align with existing conceptualizations of interpersonal incivility and malicious word-of-mouth (Cho and Kwon 2015; Golf-Papez and Veer 2022; Hornik et al. 2019), flame-bait firestorms and toxic debunking capture hybrid dynamics in which influencers actively shape both the targets and moral framing of aggression. This distinction extends current typologies by highlighting how authority and audience alignment reconfigure the social meaning of toxic participation.

Relatedly, the mechanisms of misinformation legitimation and community enmeshment advance conceptual work on influencer self-presentation and audience bonding. Existing research frames credibility signaling and controversy management as strategies for sustaining engagement and visibility (Cocker and Cronin 2017; Abidin 2019). The present findings theorize a critical escalation of these practices: credibility labor (Bahar and Hasan 2024) becomes a vehicle for institutionalizing misinformation, while parasocial intimacy (Reinikainen et al. 2020; Mardon et al. 2023) evolves into dense, inward-facing community structures that reward conformity and sanction dissent. In this way, influencer–follower relationships shift from dyadic attachment to collective enclosure, enabling the persistence of toxic echo chambers.

Third, the study contributes to theories of misinformation processing by challenging cognition-centric explanations of belief formation. Building on Kelman (1958) social influence framework, the findings show that compliance, identification, and internalization operate as collective alignment mechanisms, rather than as individual persuasion outcomes. Influencers' relational authority and homophilic audience composition create social pressures that prioritize loyalty and identity affirmation over factual evaluation. This insight reframes misinformation as a socially sustained practice, aligning with recent evidence that group identity and social belonging outweigh accuracy considerations in shaping engagement and belief (Van Bavel et al. 2024). Consequently, the persistence of toxicity, even following correction or debunking, can be understood as a function of social reinforcement rather than informational deficit.

8.2 | Managerial Implications

This research provides actionable guidance for managing influencer-driven misinformation and toxic engagement by organizing interventions around the three main actors operating in the misinformation ecosystem (Table 6): Publishers

TABLE 6 | Managerial implications.

P-area	Primary actor(s)	Key risk	Recommended managerial actions	Expected outcome
Publishers	SMIs	Influencer-generated misinformation triggers faster, denser toxic spirals due to credibility and parasocial loyalty	• Monitor high-engagement influencer content on controversial topics. • Use influencer-led corrections when misinformation is detected to control parasocial trust. • Encourage transparency and factual framing in influencer content.	Reduced escalation at the source; containment of downstream toxicity
	Brands managing influencer partnerships	Reputational spillover from influencer misconduct	• Track influencer mentions of brand names for misinformation signals. • Prepare rapid-response protocols for misinformation crises (even if influencer is not a partner). • Publicly clarify misinformation using nonadversarial tone to avoid escalation.	Lower brand exposure to toxic amplification
Platforms	Social media platforms	Algorithmic amplification of outrage-driven engagement	• Implement early-warning systems for engagement spikes tied to misinformation. • Deprioritize algorithmic amplification of toxic engagement signals. • Flag and slow diffusion of high-velocity misinformation mentioning brands.	Slower diffusion and reduced visibility of toxic content
	Brands (owned and paid media spaces)	Delayed or misaligned responses intensify backlash	• Maintain crisis playbooks tailored to toxicity types (e.g., flame-bait firestorms vs. toxic debunking). • Apply factual, dialogic moderation in comment threads.	More effective reputational containment
	Platforms and brands jointly	Monetization incentives reward toxicity	• Advocate for alternative metrics (verified reach, content quality). • Align ad spend with credibility-based signals.	Structural reduction in toxic incentives
	Social media users/ followers	Parasocial bonds and group identity reinforce toxic engagement	• Use dialogic, empathy-based communication to de-escalate hostility.	De-escalation and reduced defensiveness
People	Communities and opinion leaders	Echo-chamber reinforcement and toxic debunking	• Engage credible intermediaries (experts, micro-influencers). • Signal clear community norms and expectations.	Weakened polarization and norm-guided engagement
Cross-Ps	All actors	Different toxicity forms require different responses	Match response to toxicity type (e.g., crisis comms vs. limited attention).	More efficient and targeted interventions

(influencers), Platforms (social media platforms and brands), and People (social media users) (Johar 2025). Our findings show that misinformation originating from influencers generates more intense and qualitatively different toxicity than misinformation shared by regular users, highlighting the need for targeted, actor-specific responses.

Publishers. Because influencers function as high-credibility publishers whose parasocial bonds accelerate toxic escalation, early intervention at the source is critical. Brands should prioritize monitoring high-engagement influencer content on controversial topics and implement governance mechanisms to manage risk. When misinformation occurs, influencer-led corrections can be more effective than external condemnation, as they leverage parasocial trust and reduce follower backlash.

Platforms. Platforms shape the visibility and velocity of toxic engagement. Brands and platforms should collaborate on early-warning systems that detect engagement spikes tied to influencer misinformation and brand mentions. Given that engagement-based monetization often rewards outrage, algorithmic and incentive-level adjustments, such as deprioritizing high-velocity toxic content and favoring verified reach or credibility signals, can help contain amplification. Differentiated response infrastructures are necessary to address distinct toxicity forms.

People. Users actively drive escalation through flame-bait firestorms, toxic debunking, and C2C conflict, often reinforced by parasocial identification. Effective interventions should therefore target relational dynamics rather than individual posts. Dialogic, nonconfrontational communication and the use of credible intermediaries (e.g., experts or micro-influencers) can diffuse polarization and weaken echo-chamber effects.

Overall, coordinated action across the three Ps enables brands to mitigate reputational harm and reduce toxic engagement by investing in credibility capital, relational governance, and systemic collaboration.

9 | Limitations and Future Research Directions

This study has several limitations that also point to opportunities for future research. First, the dataset may be incomplete. Although collection procedures were systematic, they depended on content that remained publicly accessible at the time of retrieval. As platforms expand moderation policies to restrict harmful material, future access to such content may narrow, limiting visibility into the full scope of online toxicity and its dynamics.

Second, our focus on heterogeneity and breadth meant that we did not examine the underlying psychological mechanisms or broader social factors that drive toxic reactions to brand-related misinformation. Future research could explore these processes through experimental designs investigating users' disinhibition, perceived intent to deceive, or identity threat when exposed to controversial content. Likewise, the motivations behind influencers' own misinformation-sharing behavior remain an important but unexamined question, best addressed through qualitative approaches.

Our design captures toxicity as it unfolds in authentic online settings (Van Heerde et al. 2021), yet causal mechanisms could be tested under controlled conditions. Experimental or vignette-based studies might manipulate source type (influencer vs. user), misinformation strength, or topic focus (brand vs. non-brand) to isolate effects on credibility, emotion, and toxic engagement. Relatedly, future work could examine how susceptibility to group opinions (Cascio et al. 2015) or resistance to peer influence moderates these effects, clarifying why some users escalate toxicity whereas others disengage.

Although our typology distinguishes toxicity by form rather than intensity, some categories (e.g., flame-bait firestorms and toxic debunking) appear qualitatively more escalatory or coordinated. Subsequent research could measure intensity as a separate dimension to capture variation in magnitude and impact. Finally, the study analyzed interactions between influencers and users at a single point in time, leaving the evolution of influencer behavior unexamined. Longitudinal analyses could trace how influencers' toxic practices develop and identify factors that shape these behavioral trajectories over time.

Funding

The authors received no specific funding for this work.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Abidin, C. 2019. "Victim, Rival, Bully: Influencers' Narrative Cultures Around Cyberbullying." In *Narratives in Research and Interventions on Cyberbullying Among Young People*, 199–212. Springer.

Ammann, J., A. Arbenz, G. Mack, and M. Siegrist. 2025. "Consumer Support of Policy Measures to Increase Sustainability in Food Consumption." *Food Policy* 131: 102822.

Aral, S. 2020. *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health—And How We Must Adapt*. Currency.

Aranda, A. M., K. Sele, H. Etchanchu, J. Y. Guyt, and E. Vaara. 2021. "From Big Data to Rich Theory: Integrating Critical Discourse Analysis With Structural Topic Modeling." *European Management Review* 18, no. 3: 197–214.

Audrezet, A., G. De Kerviler, and J. Guidry Moulard. 2020. "Authenticity Under Threat: When Social Media Influencers Need to go Beyond Self-Presentation." *Journal of Business Research* 117: 557–569.

Avalle, M., N. Di Marco, G. Etta, et al. 2024. "Persistent Interaction Patterns Across Social Media Platforms and Over Time." *Nature* 628, no. 8008: 582–589.

Bacile, T. J., A. B. Elmadag, M. Okan, D. Dineva, and A. I. Rynarzewska. 2025. "Schadenfreude and Sympathy: Observer Reactions to Malicious Joy During Social Media Service Recovery." *Journal of Interactive Marketing* 60: 44–64.

Baele, S., L. Brace, and D. Ging. 2024. "A Diachronic Cross-Platforms Analysis of Violent Extremist Language in the Incel Online Ecosystem." *Terrorism and Political Violence* 36, no. 3: 382–405.

Bahar, V. S., and M. Hasan. 2024. "# Fakefamous: How Do Influencers Use Disinformation to Establish Long-Term Credibility on Social Media?" *Information Technology & People* 38, no. 6: 2441–2476.

Barari, M. 2023. "Unveiling the Dark Side of Influencer Marketing: How Social Media Influencers (Human vs Virtual) Diminish Followers' Well-Being." *Marketing Intelligence & Planning* 41, no. 8: 1162–1177.

Bassignana, E., V. Basile, and V. Patti. 2018. "Hurtlex: A Multilingual Lexicon of Words to Hurt." In *CEUR Workshop proceedings (vol. 2253)*, 1–6. CEUR-WS.

Belanche, D., L. V. Casaló, M. Flavián, and S. Ibáñez-Sánchez. 2021. "Understanding Influencer Marketing: The Role of Congruence Between Influencers, Products and Consumers." *Journal of Business Research* 132: 186–195.

Bonini, T., A. Caliandro, and A. Massarelli. 2016. "Understanding the Value of Networked Publics in Radio: Employing Digital Methods and Social Network Analysis to Understand the Twitter Publics of Two Italian National Radio Stations." *Information, Communication & Society* 19, no. 1: 40–58.

Bovet, A., and H. A. Makse. 2019. "Influence of Fake News in Twitter During the 2016 US Presidential Election." *Nature Communications* 10, no. 1: 7.

Brady, W. J., K. L. McLoughlin, M. P. Torres, K. F. Luo, M. Gendron, and M. J. Crockett. 2023. "Overperception of Moral Outrage in Online Social Networks Inflates Beliefs About Intergroup Hostility." *Nature Human Behaviour* 7: 917–927.

Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3, no. 2: 77–101.

Brennen, J. S., F. M. Simon, and R. K. Nielsen. 2021. "Beyond (Mis) Representation: Visuals in COVID-19 Misinformation." *International Journal of Press/Politics* 26, no. 1: 277–299.

Breves, P. L., N. Liebers, M. Abt, and A. Kunze. 2019. "The Perceived Fit Between Instagram Influencers and the Endorsed Brand: How Influencer–Brand Fit Affects Source Credibility and Persuasive Effectiveness." *Journal of Advertising Research* 59, no. 4: 440–454.

Caliandro, A., and A. Gandini. 2016. *Qualitative Research in Digital Environments: A Research Toolkit*. Routledge.

Campbell, C., and J. R. Farrell. 2020. "More Than Meets the Eye: The Functional Components Underlying Influencer Marketing." *Business Horizons* 63, no. 4: 469–479.

Cascio, C. N., M. B. O'Donnell, J. Bayer, F. J. Tinney, and E. B. Falk. 2015. "Neural Correlates of Susceptibility to Group Opinions in Online Word-of-Mouth Recommendations." *Journal of Marketing Research* 52, no. 4: 4.

CCDH. 2021. "The Disinformation Dozen." https://counterhate.com/research/the-disinformation-dozen/.

Cho, D., and K. H. Kwon. 2015. "The Impacts of Identity Verification and Disclosure of Social Cues on Flaming in Online User Comments." *Computers in Human Behavior* 51: 363–372.

Chung, S., and H. Cho. 2017. "Fostering Parasocial Relationships With Celebrities on Social Media: Implications for Celebrity Endorsement." *Psychology & Marketing* 34, no. 4: 481–495.

Cialdini, R. B., and N. J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55, no. 1: 591–621.

Cinelli, M., A. Pelicon, I. Mozetič, W. Quattrociocchi, P. K. Novak, and F. Zollo. 2021. "Dynamics of Online Hate and Misinformation." *Scientific Reports* 11, no. 1: 22083.

Coates, A. E., C. A. Hardman, J. C. G. Halford, P. Christiansen, and E. J. Boyland. 2019. "Social Media Influencer Marketing and Children's Food Intake: A Randomized Trial." *Pediatrics* 143, no. 4: e20182554.

Cocker, H. L., and J. Cronin. 2017. "Charismatic Authority and the Youtuber: Unpacking the New Cults of Personality." *Marketing Theory* 17, no. 4: 455–472.

Colliander, J. 2019. ""This Is Fake News": Investigating the Role of Conformity to Other Users' Views When Commenting on and Spreading Disinformation in Social Media." *Computers in Human Behavior* 97: 202–215.

Cova, B., and D. Dalli. 2009. "Working Consumers: The Next Step in Marketing Theory?" *Marketing Theory* 9, no. 3: 315–339.

Del Vicario, M., A. Bessi, F. Zollo, et al. 2016. "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences* 113, no. 3: 554–559.

Denny, M. J., and A. Spirling. 2018. "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do About It." *Political Analysis* 26, no. 2: 168–189.

Dessart, L., C. Veloutsou, and A. Morgan-Thomas. 2020. "Brand Negativity: A Relational Perspective on Anti-Brand Community Participation." *European Journal of Marketing* 54, no. 7: 1761–1785.

Di Domenico, G., and Y. Ding. 2023. "Between Brand Attacks and Broader Narratives: How Direct and Indirect Misinformation Erode Consumer Trust." *Current Opinion in Psychology* 54: 101716.

Di Domenico, G., D. Nunan, and V. Pitardi. 2022. "Marketplaces of Misinformation: A Study of How Vaccine Misinformation Is Legitimized on Social Media." *Journal of Public Policy & Marketing* 41, no. 4: 319–335.

Di Domenico, G. D., J. Sit, A. Ishizaka, and D. Nunan. 2021. "Fake News, Social Media and Marketing: A Systematic Review." *Journal of Business Research* 124: 329–341.

Dineva, D. 2023. "Consumer Incivility in Virtual Spaces: Implications for Interactive Marketing Research and Practice." In *The Palgrave Handbook of Interactive Marketing*, 917–937. Springer International Publishing.

Dineva, D., and J. Breitsohl. 2022. "Managing Trolling in Online Communities: An Organizational Perspective." *Internet Research* 32, no. 1: 292–311.

Dineva, D., J. Breitsohl, B. Garrod, and P. Megicks. 2020. "Consumer Responses to Conflict-Management Strategies on Non-Profit Social Media Fan Pages." *Journal of Interactive Marketing* 52: 118–136.

Dineva, D., and K. L. Daunt. 2023. "Reframing Online Brand Community Management: Consumer Conflicts, Their Consequences and Moderation." *European Journal of Marketing* 57, no. 10: 2653–2682.

Edmondson, A. C., and S. E. McManus. 2007. "Methodological Fit in Management Field Research." *Academy of Management Review* 32, no. 4: 1246–1264.

Ekinci, Y., S. Dam, and G. Buckle. 2025. "The Dark Side of Social Media Influencers: A Research Agenda for Analysing Deceptive Practices and Regulatory Challenges." *Psychology & Marketing* 42, no. 4: 1201–1214.

Erdmann, A., R. Arilla, and J. M. Ponzoa. 2022. "Search Engine Optimization: The Long-Term Strategy of Keyword Choice." *Journal of Business Research* 144: 650–662.

Escalas, J. E., and J. R. Bettman. 2005. "Self-Construal, Reference Groups, and Brand Meaning." *Journal of Consumer Research* 32, no. 3: 378–389.

Ewing, M. T., P. E. Wagstaff, and I. H. Powell. 2013. "Brand Rivalry and Community Conflict." *Journal of Business Research* 66, no. 1: 4–12.

Feng, Y., and H. J. Kim. 2024. "Why Do People Generate Toxic Speech Toward Woke Advertising? The Role of Persuasion Knowledge and Cognitive Dissonance." *Journal of Current Issues & Research in Advertising* 46, no. 1: 69–89.

Fereday, J., and E. Muir-Cochrane. 2006. "Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development." *International Journal of Qualitative Methods* 5, no. 1: 80–92.

Fombelle, P. W., C. M. Voorhees, M. R. Jenkins, et al. 2020. "Customer Deviance: A Framework, Prevention Strategies, and Opportunities for Future Research." *Journal of Business Research* 116: 387–400.

Fong, A., J. Roozenbeek, D. Goldwert, S. Rathje, and S. Van Der Linden. 2021. "The Language of Conspiracy: A Psychological Analysis of Speech Used by Conspiracy Theorists and Their Followers on Twitter." *Group Processes & Intergroup Relations* 24, no. 4: 606–623.

Garibay, I., A. V. Mantzaris, A. Rajabi, and C. E. Taylor. 2019. "Polarization in Social Media Assists Influencers to Become More Influential: Analysis and Two Inoculation Strategies." *Scientific Reports* 9, no. 1: 18592.

Garimella, K., G. D. F. Morales, A. Gionis, and M. Mathioudakis. 2018. "Quantifying Controversy on Social Media." *ACM Transactions on Social Computing* 1, no. 1: 1–27.

Gensler, S., F. Völckner, Y. Liu-Thompkins, and C. Wiertz. 2013. "Managing Brands in the Social Media Environment." *Journal of Interactive Marketing* 27, no. 4: 242–256.

Golder, P. N., M. G. Dekimpe, J. T. An, H. J. Van Heerde, D. S. U. Kim, and J. W. Alba. 2023. "Learning From Data: An Empirics-First Approach to Relevant Knowledge Generation." *Journal of Marketing* 87, no. 3: 319–336.

Golf-Papez, M., and E. Veer. 2022. "Feeding the Trolling: Understanding and Mitigating Online Trolling Behavior as an Unintended Consequence." *Journal of Interactive Marketing* 57, no. 1: 90–114.

Greve, H. R., H. Rao, P. Vicinanza, and E. Y. Zhou. 2022. "Online Conspiracy Groups: Micro-Bloggers, Bots, and Coronavirus Conspiracy Talk on Twitter." *American Sociological Review* 87, no. 6: 919–949.

Grinberg, N., K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer. 2019. "Fake News on Twitter During the 2016 US Presidential Election." *Science* 363, no. 6425: 374–378.

Guldemond, P., A. Casas Salleras, and M. Van der Velden. 2022. "Fueling Toxicity? Studying Deceitful Opinion Leaders and Behavioral Changes of Their Followers." *Politics and Governance* 10, no. 4: 336–348.

Gurrieri, L., J. Drenten, and C. Abidin. 2023. "Symbiosis or Parasitism? A Framework for Advancing Interdisciplinary and Socio-Cultural Perspectives in Influencer Marketing." *Journal of Marketing Management* 39, no. 11–12: 911–932.

Hall, R. 2025. "Chinese Manufacturers Flood TikTok With Videos Urging Americans to Buy Direct After Trump's Tariffs." *The Independent*. https://www.independent.co.uk/news/world/americas/us-politics/china-trump-tariffs-tiktok-instagram-lululemon-louis-vuitton-b2734819.html.

Han, J., and G. Balabanis. 2024. "Meta-Analysis of Social Media Influencer Impact: Key Antecedents and Theoretical Foundations." *Psychology & Marketing* 41, no. 2: 394–426.

Hanu, L., and Unitary Team. 2020. "Detoxify." https://github.com/unitaryai/detoxify.

Harff, D., C. Bollen, and D. Schmuck. 2022. "Responses to Social Media Influencers' Misinformation About COVID-19: A Pre-Registered Multiple-Exposure Experiment." *Media Psychology* 25, no. 6: 831–850.

Harrison-Walker, L. J., and Y. Jiang. 2023. "Suspicion of Online Product Reviews as Fake: Cues and Consequences." *Journal of Business Research* 160: 113780.

Van Heerde, H. J., C. Moorman, C. P. Moreau, and R. W. Palmatier. 2021. "Reality Check: Infusing Ecological Value Into Academic Marketing Research." Journal of Marketing 85, no. 2: 1–13.

Herhausen, D., S. Ludwig, D. Grewal, J. Wulf, and M. Schoegel. 2019. "Detecting, Preventing, and Mitigating Online Firestorms in Brand Communities." *Journal of Marketing* 83, no. 3: 1–21.

Holt, D., and D. Cameron. 2010. *Cultural Strategy: Using Innovative Ideologies to Build Breakthrough Brands*. OUP Oxford.

Hornik, J., R. Shaanan Satchi, and M. Rachamim. 2019. "The Joy of Pain: A Gloating Account of Negative Electronic Word-of-Mouth Communication Following an Organizational Setback." *Internet Research* 29, no. 1: 82–103.

Horton, D., and R. Richard Wohl. 1956. "Mass Communication and Para-Social Interaction: Observations on Intimacy at a Distance." *Psychiatry* 19, no. 3: 215–229.

Hotz-Behofsits, C., N. Wlömert, and N. Abou Nabout. 2025. "Natural Affect Detection (NADE): Using Emojis to Infer Emotions From Text." *Journal of Marketing*: 00222429251315088.

Hovland, C. I., and W. Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *Public Opinion Quarterly* 15, no. 4: 635–650.

Hughes, C., V. Swaminathan, and G. Brooks. 2019. "Driving Brand Engagement Through Online Social Influencers: An Empirical Investigation of Sponsored Blogging Campaigns." *Journal of Marketing* 83, no. 5: 78–96.

Humphreys, A., and R. J. H. Wang. 2018. "Automated Text Analysis for Consumer Research." *Journal of Consumer Research* 44, no. 6: 1274–1306.

Husemann, K. C., F. Ladstaetter, and M. K. Luedicke. 2015. "Conflict Culture and Conflict Management in Consumption Communities." *Psychology & Marketing* 32, no. 3: 265–284.

Influencer Marketing Hub. 2023. "The State of Influencer Marketing: Benchmark Report 2023 [E-Book]." https://influencermarketinghub.com/ebooks/Influencer_Marketing_Benchmark_Report_2023.pdf.

Johar, G. V. 2022. "Untangling the Web of Misinformation and False Beliefs." *Journal of Consumer Psychology* 32, no. 2: 374–383.

Johar, G. V. 2025. "Call for Papers | Journal of Public Policy & Marketing: Mitigating Misinformation." American Marketing Association. https://www.ama.org/2025/01/06/call-for-papers-journal-of-public-policy-marketing-mitigating-misinformation/.

Kapitan, S., and D. H. Silvera. 2016. "From Digital Media Influencers to Celebrity Endorsers: Attributions Drive Endorser Effectiveness." *Marketing Letters* 27: 553–567.

Karagür, Z., J. M. Becker, K. Klein, and A. Edeling. 2022. "How, Why, and When Disclosure Type Matters for Influencer Marketing." *International Journal of Research in Marketing* 39, no. 2: 313–335.

Kelman, H. C. 1958. "Compliance, Identification, and Internalization Three Processes of Attitude Change." *Journal of Conflict Resolution* 2, no. 1: 51–60.

Kelman, H. C. 2006. "Interests, Relationships, Identities: Three Central Issues for Individuals and Groups in Negotiating Their Social Environment." *Annual Review of Psychology* 57, no. 1: 1–26.

Kim, J. W., A. Guess, B. Nyhan, and J. Reifler. 2021. "The Distorting Prism of Social Media: How Self-Selection and Exposure to Incivility Fuel Online Comment Toxicity." *Journal of Communication* 71, no. 6: 922–946.

Koorank Beheshti, M., M. Gopinath, S. Ashouri, and S. Zal. 2023. "Does Polarizing Personality Matter in Influencer Marketing? Evidence From Instagram." *Journal of Business Research* 160: 113804.

Labrecque, L. I. 2014. "Fostering Consumer–Brand Relationships in Social Media Environments: The Role of Parasocial Interaction." *Journal of interactive marketing* 28, no. 2: 134–148.

Leung, F. F., F. F. Gu, and R. W. Palmatier. 2022. "Online Influencer Marketing." *Journal of the Academy of Marketing Science* 50: 226–251.

Liao, J., Y. Ye, R. Filieri, P. Du, and Y. Jiang. 2024. "Why Did You Delete My Comment? Investigating Observing Consumers' Reactions to Comment-Deletion-Clues During a Brand Crisis." *Psychology & Marketing* 41, no. 10: 2478–2492.

Lin, Y. W., S. Yang, W. Han, and J. G. Lu. 2024. "The Black Lives Matter Movement Mitigates Bias Against Racial Minority Actors." *Proceedings of the National Academy of Sciences* 121, no. 29: e2307726121.

Literat, I., and N. Kligler-Vilenchik. 2021. "How Popular Culture Prompts Youth Collective Political Expression and Cross-Cutting Political Talk on Social Media: A Cross-Platform Analysis." *Social Media+ Society* 7, no. 2: 20563051211008821.

Lou, C. 2022. "Social Media Influencers and Followers: Theorization of a Trans-Parasocial Relation and Explication of Its Implications for Influencer Advertising." *Journal of Advertising* 51, no. 1: 4–21.

Luedicke, M. K., C. J. Thompson, and M. Giesler. 2010. "Consumer Identity Work as Moral Protagonism: How Myth and Ideology Animate a Brand-Mediated Moral Conflict." *Journal of Consumer Research* 36, no. 6: 1016–1032.

Lunardo, R., M. Alemany Oliver, and S. Shepherd. 2023. "How Believing in Brand Conspiracies Shapes Relationships With Brands." *Journal of Business Research* 159: 113729.

Mangiò, F., and G. Di Domenico. 2022. "All That Glitters Is Not Real Affiliation: How to Handle Affiliate Marketing Programs in the Era of Falsity." *Business Horizons* 65, no. 6: 765–776.

Mardon, R., H. Cocker, and K. Daunt. 2023. "How Social Media Influencers Impact Consumer Collectives: An Embeddedness Perspective." *Journal of Consumer Research* 50: 617–644.

Mardon, R., M. Molesworth, and G. Grigore. 2018. "Youtube Beauty Gurus and the Emotional Labour of Tribal Entrepreneurship." *Journal of Business Research* 92: 443–454.

Martel, C., J. Allen, G. Pennycook, and D. G. Rand. 2024. "Crowds Can Effectively Identify Misinformation at Scale." *Perspectives on Psychological Science* 19, no. 2: 477–488.

Marwick, A., and D. Boyd. 2011. "To See and Be Seen: Celebrity Practice on Twitter." *Convergence: The International Journal of Research Into New Media Technologies* 17, no. 2: 139–158.

McCormick, K. 2016. "Celebrity Endorsements: Influence of a Product-Endorser Match on Millennials Attitudes and Purchase Intentions." *Journal of Retailing and Consumer Services* 32: 39–45.

Metzger, M. J., A. J. Flanagin, P. Mena, S. Jiang, and C. Wilson. 2021. "From Dark to Light: The Many Shades of Sharing Misinformation Online." *Media and Communication* 9, no. 1: 134–143

Mills, A. J., and K. Robson. 2020. "Brand Management in the Era of Fake News: Narrative Response as a Strategy to Insulate Brand Value." *Journal of Product & Brand Management* 29, no. 2: 159–167.

Milne, M. J., and R. W. Adler. 1999. "Exploring the Reliability of Social and Environmental Disclosures Content Analysis." *Accounting, Auditing & Accountability Journal* 12, no. 2: 237–256.

Miranda, S., N. Berente, S. Seidel, H. Safadi, and A. Burton-Jones. 2022. "Editor's Comments: Computationally Intensive Theory Construction: A Primer for Authors and Reviewers." *MIS Quarterly* 46, no. 2: iii–xviii.

Monahan, L., J. A. Espinosa, J. Langenderfer, and D. J. Ortinau. 2023. "Did You Hear Our Brand Is Hated? The Unexpected Upside of Hate-Acknowledging Advertising for Polarizing Brands." *Journal of Business Research* 154: 113283.

Mulcahy, R., R. Barnes, R. de Villiers Scheepers, S. Kay, and E. List. 2024. "Going Viral: Sharing of Misinformation by Social Media Influencers." *Australasian Marketing Journal* 33, no. 3: 296–309.

Nepomuceno, M. V., H. Rahemi, T. Cenesizoglu, and L. Charlin. 2023. "Should We Feed the Trolls? Using Marketer-Generated Content to Explain Average Toxicity and Product Usage." *Journal of Interactive Marketing* 58, no. 4: 440–462.

Ohanian, R. 1990. "Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness." *Journal of advertising* 19, no. 3: 39–52.

Ooms, J. 2023. "_cld3: Google's Compact Language Detector 3_." R Package Version 1.6.0. https://CRAN.R-project.org/package=cld3.

Papakyriakopoulos, O., and E. Goodman. 2022 April. "The Impact of Twitter Labels on Misinformation Spread and User Engagement: Lessons From Trump's Election Tweets." In *Proceedings of the ACM Web Conference 2022*, 2541–2551.

Pennycook, G., and D. G. Rand. 2021. "The Psychology of Fake News." *Trends in Cognitive Sciences* 25, no. 5: 388–402.

Perspective AI. 2024. "Google Project Jigsaw." Perspective. https://www.perspectiveapi.com/#/.

Rao, H., and H. R. Greve. 2024. "The Plot Thickens: A Sociology of Conspiracy Theories." *Annual Review of Sociology* 50: 191–207.

Reinikainen, H., J. Munnukka, D. Maity, and V. Luoma-Aho. 2020. "'You Really Are a Great Big Sister'–Parasocial Relationships, Credibility, and the Moderating Role of Audience Comments in Influencer Marketing." *Journal of Marketing Management* 36, no. 3–4: 279–298.

Rinker, T. W. 2018. "textclean: Text Cleaning Tools Version 0.9.3." Buffalo, New York. https://github.com/trinker/textclean.

Romani, S., S. Grappi, L. Zarantonello, and R. P. Bagozzi. 2015. "The Revenge of the Consumer! How Brand Moral Violations Lead to Consumer Anti-Brand Activism." *Journal of brand Management* 22: 658–672.

Scheibenzuber, C., L. M. Neagu, S. Ruseti, et al. 2023. "Dialog in the Echo Chamber: Fake News Framing Predicts Emotion, Argumentation and Dialogic Social Knowledge Building in Subsequent Online Discussions." *Computers in Human Behavior* 140: 107587.

Schmidt, A. L., A. Peruzzi, A. Scala, et al. 2020. "Measuring Social Response to Different Journalistic Techniques on Facebook." *Humanities and Social Sciences Communications* 7, no. 1: 17.

Schoenmueller, V., O. Netzer, and F. Stahl. 2022. "Polarized America: From Political Partisanship to Preference Partisanship." *Marketing Science Frontiers Forthcoming*. https://ssrn.com/abstract=3471477.

Scholz, J. 2021. "How Consumers Consume Social Media Influence." *Journal of Advertising* 50, no. 5: 510–527.

Scholz, J., and A. N. Smith. 2019. "Branding in the Age of Social Media Firestorms: How to Create Brand Value by Fighting Back Online." *Journal of Marketing Management* 35, no. 11–12: 1100–1134.

Shahbaznezhad, H., R. Dolan, and M. Rashidirad. 2021. "The Role of Social Media Content Format and Platform in Users' Engagement Behavior." *Journal of Interactive Marketing* 53: 47–65.

Shao, C., G. L. Ciampaglia, O. Varol, K. C. Yang, A. Flammini, and F. Menczer. 2018. "The Spread of Low-Credibility Content by Social Bots." *Nature Communications* 9, no. 1: 4787.

Shehzala, A. K. Jaiswal, V. Vemireddy, and F. Angeli. 2024. "Social Media 'Stars' vs 'The Ordinary' Me: Influencer Marketing and the Role of Self-Discrepancies, Perceived Homophily, Authenticity, Self-Acceptance and Mindfulness." *European Journal of Marketing* 58, no. 2: 590–631.

Statista. 2023. "Influencer Marketing Market Size Worldwide From 2016 to 2023 (in billion U.S. Dollars)." https://www-statista-com/statistics/1092819/global-influencer-market-size/.

Stewart, N. K., A. Al-Rawi, C. Celestini, and N. Worku. 2023. "Hate Influencers' Mediation of Hate on Telegram:' We Declare War Against the Anti-White System'." *Social Media+ Society* 9, no. 2: 20563051231177915.

Stieglitz, S., M. Mirbabaie, T. Kroll, and J. Marx. 2019. "'Silence' as a Strategy During a Corporate Crisis—The Case of Volkswagen's 'Dieselgate.'" *Internet Research* 29, no. 4: 921–939.

Suler, J. 2004. "The Online Disinhibition Effect." *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society* 7, no. 3: 321–326.

Thomas, V. L., K. Fowler, and F. Taheran. 2024. "How Social Media Influencer Collaborations Are Perceived by Consumers." *Psychology & Marketing* 41, no. 1: 168–183.

Timothy Coombs, W., and S. Jean Holladay. 2014. "How Publics React to Crisis Communication Efforts: Comparing Crisis Response Reactions Across Sub-Arenas." *Journal of Communication Management* 18, no. 1: 40–57.

Trope, Y., and N. Liberman. 2010. "Construal-Level Theory of Psychological Distance." *Psychological Review* 117, no. 2: 440–463.

Valsesia, F., D. Proserpio, and J. C. Nunes. 2020. "The Positive Effect of Not Following Others on Social Media." *Journal of Marketing Research* 57, no. 6: 1152–1168.

Van Bavel, J. J., S. Rathje, M. Vlasceanu, and C. Pretus. 2024. "Updating the Identity-Based Model of Belief: From False Belief to the Spread of Misinformation." *Current Opinion in Psychology* 56: 101787.

Varol, O., E. Ferrara, C. Davis, F. Menczer, and A. Flammini. 2017. "Online Human-Bot Interactions: Detection, Estimation, and Characterization." *Proceedings of the International AAAI Conference on Web and Social Media* 11, no. 1: 280–289.

Visentin, M., G. Pizzi, and M. Pichierri. 2019. "Fake News, Real Problems for Brands: The Impact of Content Truthfulness and Source Credibility on Consumers' Behavioral Intentions Toward the Advertised Brands." *Journal of Interactive Marketing* 45, no. 1: 99–112.

Vogels, E. A. 2021. "The State of Online Harassment." *Pew Research Center* 13: 625. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/.

Vosoughi, S., D. Roy, and S. Aral. 2018. "The Spread of True and False News Online." *Science* 359, no. 6380: 1146–1151.

Walker, S., D. Mercea, and M. Bastos. 2019. "The Disinformation Landscape and the Lockdown of Social Platforms." *Information, Communication & Society* 22, no. 11: 1531–1543.

Wallace, E., and I. Buil. 2025. "Individual Differences in Perceiving Disinformation Sharing as Opinion Leadership: Effects of Dark Triad Traits, Need for Uniqueness, and Green Identity." *Personality and Individual Differences* 238: 113105.

Yan, X., J. Guo, Y. Lan, and X. Cheng. 2013. "A Biterm Topic Model for Short Texts." In *Proceedings of the 22nd International Conference on World Wide Web*, 1445–1456.

Yuan, S., and C. Lou. 2020. "How Social Media Influencers Foster Relationships With Followers: The Roles of Source Credibility and Fairness in Parasocial Relationship and Product Interest." *Journal of Interactive Advertising* 20, no. 2: 133–147.

Zhu, L., and Y. Wang. 2025. "Acting Real: A Cross-Cultural Investigation of Finfluencer Strategic Authenticity." *International Journal of Advertising* 44, no. 1: 164–183.

Zoizner, A., and A. Levy. 2025. "How Social Media Users Adopt the Toxic Behaviors of Ingroup and Outgroup Accounts." *Journal of Computer-Mediated Communication* 30, no. 6: zmaf018.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.
Appendix.