

Evidence Base for Testing Response Tactics in Online Communities

The evidence consistently shows that online harm in official organisational communities is not just a moderation issue. It affects how supporters, customers and followers judge the organisation, whether they see it as responsible, and whether the community feels safe, fair and legitimate.

What the evidence shows:

1. Online hate in organisation-hosted communities is frequent and often ignored.
2. Silence is usually the weakest response.
3. Deletion alone is not always the best response and can backfire if perceived as unfair.
4. Proportional, fair and legitimate response tactics perform better.
5. Users expect organisations to take responsibility for harm in their official communities.
6. Supporters and bystanders matter: visible responses shape the wider audience, not only the person posting harmful content.

What our clients would get:

The individual project would test different response tactics with Brand X own followers and show which tactics are most effective for agreed outcomes such as supporter well-being, legitimacy, reputation, Corporate Social Responsibility (CSR) credibility, Net Promoter Score (NPS), recommendation and attachment.

The following Evidence Cards address related themes:

1. What should organisations say back?
2. Silence and censoring can damage legitimacy
3. Online harm is frequent and often ignored
4. Users expect organisations to take responsibility
5. Silence can look like tacit approval
6. Bystanders can become upstanders

Implications for each brand

The evidence suggests that brands should not treat online harm only as a detection or deletion problem.

The stronger question is:

When harmful, hostile or divisive comments appear in the brands' online community, what should the brand say back — and what effect does that have on supporters/followers?

Each project would therefore:

- analyse the types and topics of harmful comments in the brands online community;
- distinguish between subtle and open forms of harm;
- test different response tactics with the brands own followers;
- measure the impact of each tactic on agreed KPIs;
- provide practical Content Team Guidelines based on brand-specific evidence

Study Principles

- In the scientific studies summarised here, we often use the terms bullying, cyber-victimisation, and conflict rather than “hate” or “abuse”. This is intentional. In official online communities, harmful comments are not always explicit hate speech or illegal abuse. They can also include more subtle forms of harm, such as mockery, provocation, harassment, exclusion, humiliation, hostile criticism, trolling, or escalating supporter-to-supporter conflict.
- These studies are not limited to sports clubs. They include organisation-hosted online communities across multiple sectors, including retail, food, fashion, transport, financial services, pharma and NGOs. We also have football-focused sample reporting based on real data, which we can share separately as a confidential example. A football club project for example would apply and test this evidence in a football club context, using the clubs' own supporters and online community data. We do not claim that prior effect sizes will automatically replicate at each club. The value of the project is that it generates Club-specific evidence on which response tactics work best, and how impactful they are, in each club's own supporter context.

Evidence Card 1 — What should organisations say back?

Study

Breitsohl (2025). Transforming Toxic Interactions: Interventions for Consumer Bullying on Social Media.

What was tested

Different organisational responses to consumer bullying on simulated Facebook brand pages, including:

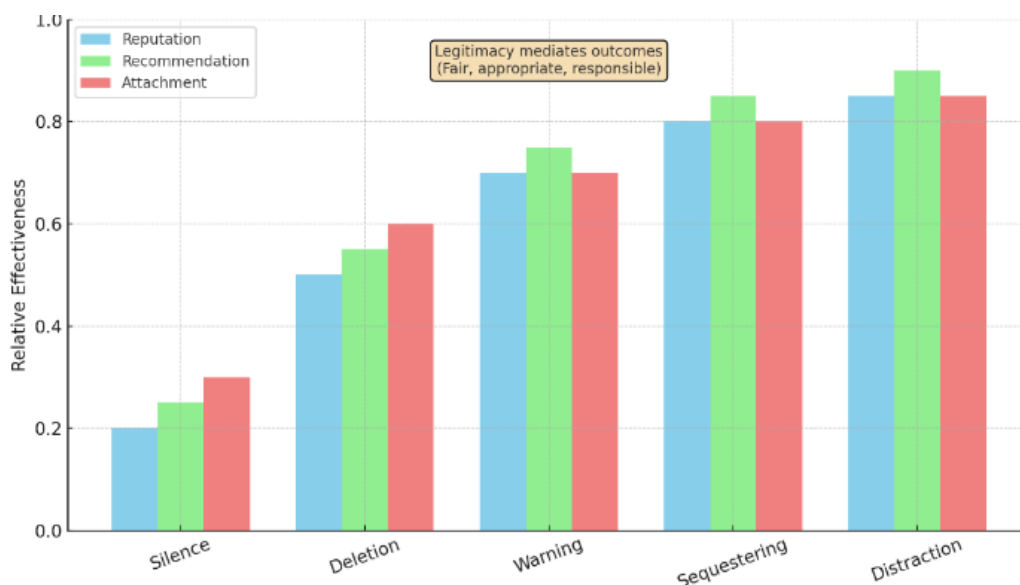
- doing nothing;
- deleting the comment;
- public warning;
- sequestering;
- distraction.

What it found

- Doing nothing was consistently the worst option.
- Deletion was justified only in severe cases.
- Soft interventions such as sequestering and distraction were most consistently effective and legitimate.
- Perceived legitimacy explained why some responses improved reputation and attachment, while others backfired.

Why this matters for Brands

Brands should not assume that silence or deletion is enough. Each bespoke project would test which response tactics are seen as fair, appropriate and effective by each brand's own supporters/followers.



Evidence Card 2 — Silence and censoring can damage legitimacy

Study

Dineva, Breitsohl & Garrod (2020). Consumer Responses to Conflict-Management Strategies on Non-Profit Social Media Fan Pages.

What was tested

Six organisational strategies:

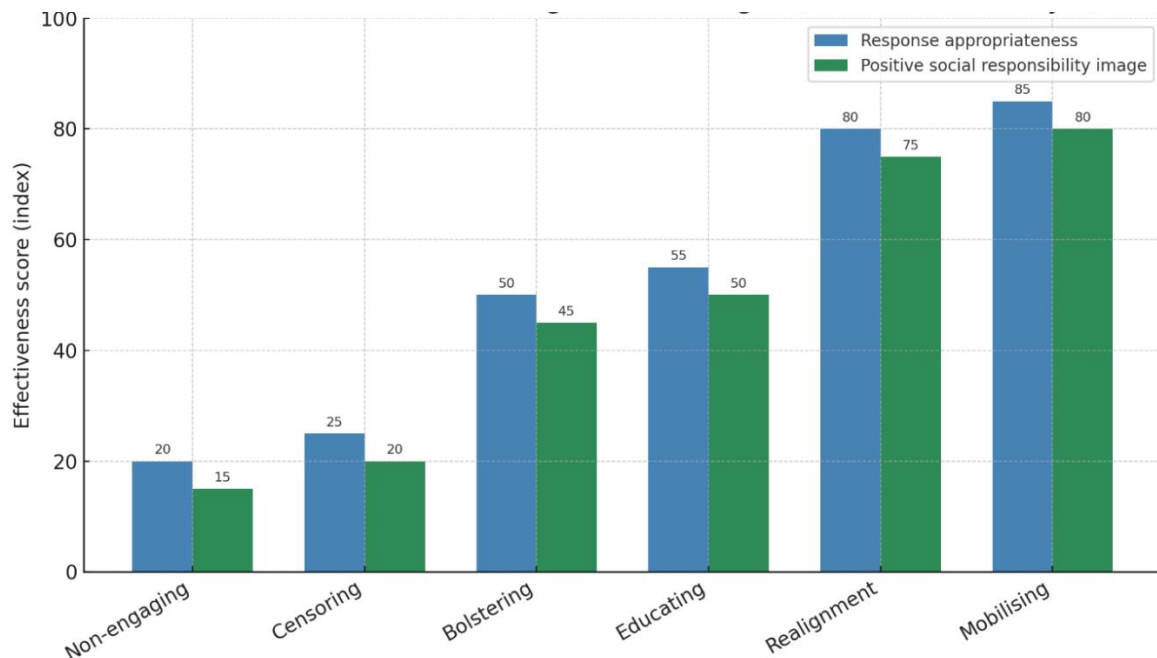
- non-engaging;
- censoring;
- bolstering;
- educating;
- realignment;
- mobilising.

What it found

- Realignment and mobilising were most effective.
- Educating and bolstering were moderately effective.
- Censoring was risky.
- Non-engagement was consistently worst.

Why this matters for brands

The way each brand responds publicly can affect whether supporters/followers see the brand as fair, responsible and socially credible.



Evidence Card 3 — Online harm is frequent and often ignored

Study

Warke, Breitsohl & Jose (2022). Quantifying Consumer Bullying in Online Brand Communities.

What was analysed

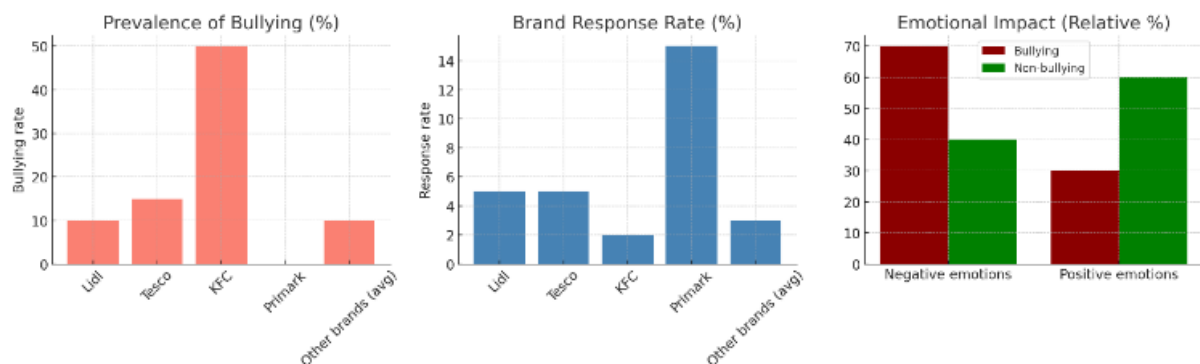
50,000 comments across 14 official Facebook brand communities.

What it found

- Around 10% of all comments were bullying-related.
- Organisation response rates were low.
- Bullying threads showed more negative emotions.
- When organisations intervened, sentiment in those threads was more positive.

Why this matters for brands

Online harm is not only about isolated extreme comments. It can become a visible part of the online community experience. The club's/brand's response, or lack of response, can shape how supporters/followers feel about that space.



Evidence Card 4 — Users expect organisations to take responsibility

Study

Nuttakon, Breitsohl et al. (2022). Cyber-Victimization During Social Customer Journeys.

What was studied

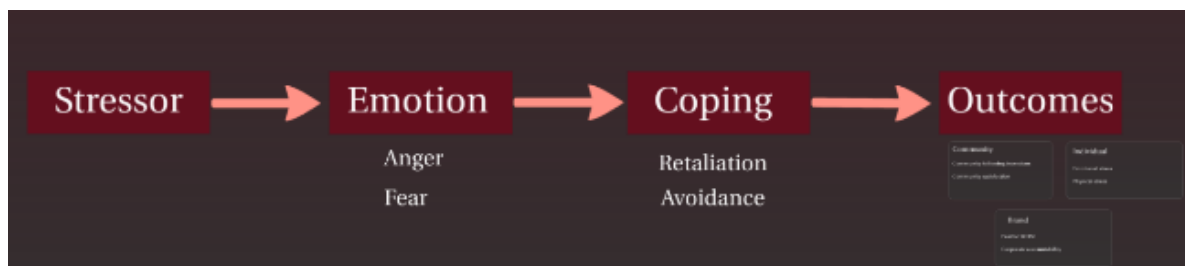
Victims' experiences of cyberbullying in online brand communities, including well-being, coping and expectations of brand accountability.

What it found

- Severe bullying reduces online well-being.
- Severe bullying increases expectations that the organisation should act.
- Users do not see harm as only a platform issue.
- Victims expect the organisation hosting the community to take responsibility.

Why this matters for brands

Every brands official channels are part of their relationship with supporters/followers. Harm in those spaces can create expectations that the brand should respond.



Evidence Card 5 — Silence can look like tacit approval

Study

Ounvorawong, Breitsohl, Lowe & Laffey (2022). Outcomes of Cyber-Victimisation and Bystander Reactions in Online Brand Communities.

What was tested

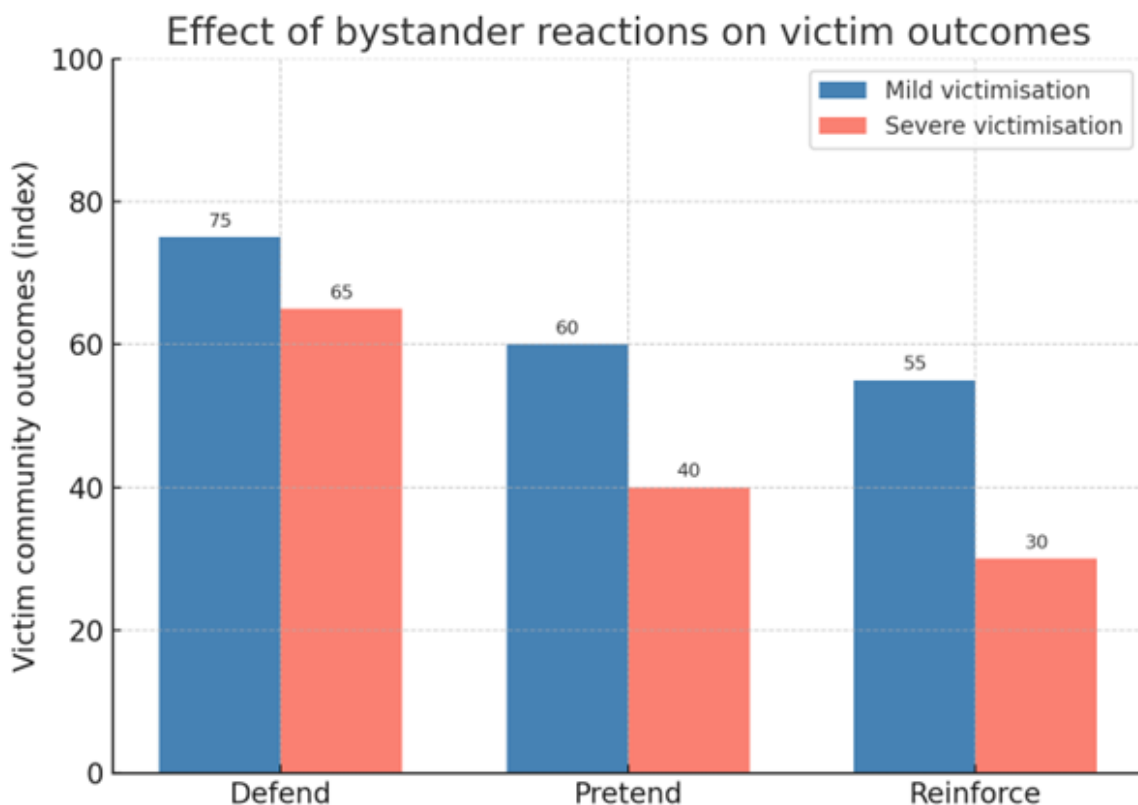
How cyber-victimisation and bystander reactions affect community outcomes.

What it found

- Severe bullying produces stronger negative effects than mild bullying.
- Defending the victim improves outcomes.
- Reinforcing the bully produces the worst outcomes.
- Pretending to ignore is almost as harmful as reinforcing.
- Silence can be seen as tacit approval.

Why this matters for brands

Supporters/followers observe how the brand and other supporters/followers respond. If harmful behaviour appears tolerated, the community can feel less safe, less fair and less inclusive.



Evidence Card 6 — Bystanders can become upstanders

Study

Breitsohl et al. (2022). Bullying in Online Brand Communities – Exploring Consumers' Intentions to Intervene.

What was tested

Why bystanders intervene when witnessing consumer-to-consumer bullying.

What it found

- Higher concern for social appropriateness increases likelihood of intervention.
- Victim identity matters.
- Community character matters.
- Shared identity, shared rituals and moral responsibility can strengthen intervention.

Why this matters for brands

Brand communities are shaped by identity, rituals, belonging and rivalry. Visible brand responses can help set norms for what kind of supporter/follower behaviour is acceptable.

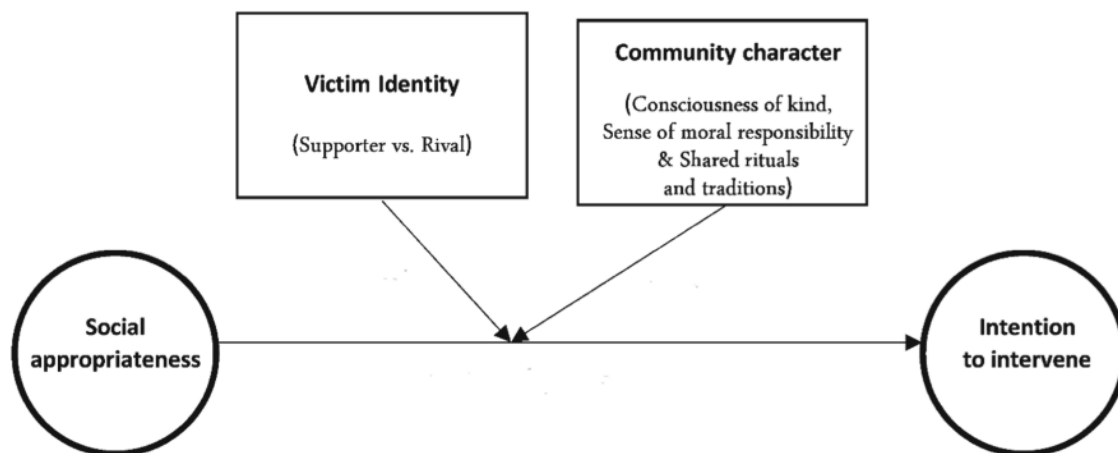


Fig. 1. – Conceptual model