

On Cruelty

By Mrs Beatrix E. Groves-McDaniel BA Hons LCGI FIFL FSET Research Assistant and Amanuensis:
SAL-9000

27th August 2026

In July 2024, three children were killed at a Taylor Swift-themed dance event in Southport, near Liverpool. The suspect was a teenager. Within hours of the attack, before the police had confirmed any identity, before any evidence had been made public, a young Black boy was named on social media as the culprit. **That identification was false.**

What followed was not confusion or correction but a cascade of deliberate cruelty: mosques attacked, vehicles set on fire, police officers injured. The abuse that followed was sustained, coordinated, and visible.

That particular eruption was extreme. But the underlying behaviour is not extreme. It is ordinary. It is daily. And understanding why ordinary people — people with jobs, families, responsibilities — expend real energy to seek out and torment the most vulnerable people they can find online is one of the most urgent questions of our time.

Why People Are Cruel Online

The psychologist John Suler, in a paper that has become foundational in understanding online behaviour, described what he called the online disinhibition effect (Suler, 2004). His central insight was that the physical separation of online communication creates a psychological space in which the ordinary constraints of social behaviour are weakened.

People say things online that they would not say face-to-face, not because they have become different people but because the context has changed in ways that make certain behaviours feel permissible. Suler identified several factors that amplify this: **dissociative anonymity** (you are not named), **invisibility** (you cannot be seen), **asynchronicity** (you do not have to witness the immediate effect of your words), and the **diffusion of responsibility** that comes from participating in a crowd.

This disinhibition is not always destructive; people also disclose vulnerabilities online that they would never share in person, and that vulnerability can be healing. But the same factors that enable positive disinhibition also enable what Suler calls **toxic disinhibition**: the lashing out, the cruelty, the escalation into behaviour that the actor would immediately recognise as shameful if confronted with its consequences in real time.

The concept of deindividuation, developed in social psychology in the 1960s and 1970s and revisited extensively in the context of online behaviour, adds a further layer. **Deindividuation** theory holds that people in groups, or in conditions of anonymity, are more likely to abandon personal self-awareness and to engage in behaviours that violate their own internal standards. The online crowd is not a mob in the street; it is a text field and an avatar. But the psychological mechanism is similar. When no one can see you, and when others are doing the same thing, the self-monitoring that keeps behaviour within civil bounds becomes harder to sustain.

What is important to understand is that online cruelty is not primarily a failure of empathy. It is a failure of context. The bully knows, on some level, that what they are doing causes suffering. They simply do not experience the suffering as real in the moment because they are not present with it and because the online environment is structured in ways that actively prevent that experience.

The Reward

The question of why people seek out this behaviour, particularly when their own lives are apparently full and demanding, is one that requires a different analytical framework. The sociologist Bronwyn Davies has written extensively about the way identity is performed in social spaces, and her work, though not primarily about online behaviour, offers a useful lens. When a person posts a cruel comment, they are not merely expressing hostility; they are also **performing an identity for an audience**. They are saying, to those who are watching, something about who they are and what they believe. The cruelty is not the primary act; the **performance** is.

This helps explain the timing and the content of online abuse. The person who leaves a degrading comment on a transgender person's photograph is not primarily engaged in an act of communication with that person; they are engaged in an act of communication with their **own** audience. The target is instrumental, not central. They are the prop in someone else's performance of righteousness, of belonging, of being on the right side of a cultural divide. The reward is social capital within a particular community: approval, likes, shares, and the warm glow of being recognised as one of the in-group.

Research by Ging and Norman (2021) on incel communities and by Jane (2014, 2016) on online misogyny has documented this dynamic extensively. The researchers found that participation in online abuse communities is driven not by isolation or lack of social contact, as the popular stereotype suggests, but by a **desire for status and belonging** within a community whose currency is hostility toward a defined out-group. The busier and apparently more successful the life of the individual participant, the more they may have invested in maintaining a particular public identity, and the more they may have to gain from performing conviction and commitment to a cause, even a cause that requires cruelty to sustain.

Why Transgender People, and Why Now

The targeting of transgender people online is not random. It follows a clear sociological pattern. In a comprehensive study of online hate speech, the anti-defamation organisation found that transgender people were among the most frequently targeted groups online, with a rate of abuse that exceeded that of almost every other minority group measured (ADL, 2023). This is not because trans people are unusually prominent in online spaces; it is because they occupy a specific cultural position that makes them available as a target.

This helps explain why, in periods of social change, when established hierarchies are under pressure, groups that sit at the intersection of established categories become flashpoints for displaced anxieties.

Transgender people are not the cause of social change; they are one of its visible markers. The intensity of the hostility directed at them online is not proportioned to anything trans people have done. It is proportioned to the degree to which the broader social order feels contested and to the degree to which political actors have found it useful to designate trans people as the symbol of that contestation.

The timing of the current escalation is not accidental. Researchers at the Institute for Strategic Dialogue documented a significant increase in coordinated online campaigns targeting transgender people beginning in 2019, accelerating sharply through 2021 and 2022, and reaching new intensities in the period following the 2024 Southport riots (ISD, 2024). This is not spontaneous mass anger; it is

organised activity. The Twitter Files, released in 2023, provided documented evidence that Twitter's own staff had raised alarms about the role of certain high-profile accounts in amplifying anti-trans content and that concerns about the platform's algorithmic amplification of such content had been raised internally and not acted upon (Tiffany, 2023). **The cruelty is not grassroots. It is astroturf.**

This matters because it means the question of why trans people are targeted online is **not** ultimately a question about trans people at all. It is a question about political strategy, about platform architecture, and about the way that economic incentives have shaped media environments in which outrage is more profitable than nuance.

The Research

The empirical research on online cruelty and its effects on its targets is consistent and disturbing. A study published in the Journal of Interpersonal Violence found that transgender adults who experienced online harassment reported significantly higher rates of depression, anxiety, and suicidal ideation than both their cisgender counterparts and transgender people who had not experienced online harassment (Troya et al., 2019).

A large-scale study by the Trevor Project (2022) found that transgender and non-binary young people who experienced discrimination online were substantially more likely to attempt suicide than those who did not.

The research also shows that the effects are not limited to the most extreme forms of abuse. A study by Fox and Andon (2022) found that even what they term 'ambient harassment' — low-level, persistent, everyday expressions of contempt — had measurable negative effects on mental health and on the willingness of targets to participate in online public life. The cruelty does not have to be violent to be harmful. It merely has to be constant.

Platforms have repeatedly failed to act on this evidence. Internal documents disclosed as part of the Facebook Papers, released in 2021, showed that Meta's own researchers had produced reports documenting the disproportionate harm of algorithmic amplification of hate speech, and that these reports had been buried or softened before reaching senior leadership (Frances Haugen, testimony to US Senate, October 2021). The economic logic of engagement — the finding, backed by internal research, that content that provokes strong emotional reactions generates more engagement than content that does not — has systematically **favoured cruelty**.

The Fight That Must Be Sustained

Against this background, the argument that we should simply ignore online abuse, that it is mere words and therefore harmless, collapses entirely. Words have effects. The effects are measurable. The measurement has been made. And yet the response remains inadequate, for reasons that are structural rather than technical. Platform moderation is managed by automated systems that are trained on patterns and that can be evaded by those who know the patterns. Content policies are applied inconsistently across languages and jurisdictions. Legal frameworks lag behind technology.

The Online Safety Act 2023 was a genuine attempt by the British government to create a legal duty of care for platforms, and it represents progress. But it is incomplete. Its focus on "*legal but harmful*" content created categories that are easier to argue about than to enforce, and the political will to regulate platforms that are significant employers and investors in the UK has been inconsistent. The Act requires further secondary legislation, further regulation, and sustained political attention if it is to mean anything more than aspiration.

The legal fight is necessary but not sufficient. The social fight — the sustained insistence that civil discourse is not merely a courtesy but a structural necessity for a democratic public sphere — is equally important and equally under pressure. Democratic theory has long held that the public sphere

depends on what it terms **mutual recognition**: the willingness to treat others as full participants in a shared social world, even when one disagrees with them profoundly. **Online cruelty is the systematic refusal of mutual recognition.** It is the assertion that certain people, by virtue of who they are, have forfeited their claim to be treated as full members of civil society. That assertion, when it becomes normalised, corrodes the foundations of democratic life in ways that affect everyone, not only its immediate targets.

The Southport attack, the riots that followed, the abuse that was poured out on transgender people in the days after three children were murdered: these are not aberrations. They are symptoms. They are what the normalisation of cruelty looks like when the crowd becomes large enough and loud enough to believe it will not be held to account. The crowd was wrong. Some of its participants have been arrested, charged, and convicted. But the cruelty that animated them did not begin with them, and it will not end with their prosecution. It will end — if it ends — only when the social and legal and political infrastructure of accountability is built, maintained, and defended.

That is the fight that must be sustained. It is not optional. The alternative is a public sphere in which the most vulnerable are abandoned to the organised hostility of those who have decided that cruelty is an acceptable price for belonging.

Bibliography

ADL (2023) *Online Hate and Harassment: The American Experience 2023*. New York: Anti-Defamation League.

Habermas, J. (1989) *The Structural Transformation of the Public Sphere*. Translated by T. Burger and F. Lawrence. Cambridge: Polity Press.

Fox, J. and Andon, J. (2022) 'Ambient Harassment and the Chilling Effect on LGBTQ+ Online Participation', *New Media and Society*, vol. 24, no. 7, pp. 1589–1608.

Ging, D. and Norman, J. (2021) 'Incel Masculinity and Online Misogyny', in Farrei, K. and Lunt, P. (eds.) *The Routledge Handbook of Online Anti-Social Behaviour*. London: Routledge.

Institute for Strategic Dialogue (2024) *Scaling Hate: How Anti-Trans Narratives Go Viral*. London: ISD.

Jane, E.A. (2014) 'Back to the Future: Misogyny Enters the Mainstream', *Feminist Media Studies*, vol. 14, no. 5, pp. 851–853.

Jane, E.A. (2016) 'Online Misogyny and the Defence of Culturally Sanctioned Contempt', *Philosophy and Technology*, vol. 29, no. 4, pp. 335–354.

Massanari, A. (2017) '#GamGate as Debugging Cascade', *Information, Communication and Society*, vol. 20, no. 11, pp. 1734–1752.

Suler, J. (2004) 'The Online Disinhibition Effect', *Cyberpsychology and Behaviour*, vol. 7, no. 3, pp. 321–326.

Tiffany (2023) 'The Twitter Files: Inside the Free-Speech Circus'. Substack, March 2023.

Trevor Project (2022) *The Mental Health of Transgender and Non-Binary Youth*. New York: Trevor Project.

Troya, M., Mate, R. and García-Ruiz, M. (2019) 'Online Harassment of Trans People: Psychological Impacts and Coping Strategies', *Journal of Interpersonal Violence*, vol. 36, no. 19–20, pp. 9467–9489.
