

The New York Times

<https://www.nytimes.com/2019/08/20/opinion/internet-harassment-restorative-justice.html>

# Could Restorative Justice Fix the Internet?

Perhaps. But it relies on people being capable of shame, so ...

Aug. 20, 2019



By Charlie Warzel

Mr. Warzel is an Opinion writer at large.

As we all spend our days yelling at one another online, it's easy to despair and wonder: Is there any way to fix our toxic internet?

Micah Loewinger, a producer for WNYC's "On the Media," was pondering this question when he met Lindsay Blackwell, a Ph.D. student at the University of Michigan who studies online harassment. Ms. Blackwell, also a researcher at Facebook, had been toying with the idea of applying the principles of the restorative justice movement to online content moderation (you can listen to their episode here).

Restorative justice is an alternative form of criminal justice that focuses on mediation. Often, an offender will meet with the victim and the broader community with a chance to make amends. The confrontation, advocates of the technique argue, helps the offender come to terms with the crime while giving the victim a chance to be heard. If the relationship is repaired and the harm to the victim reduced, the offender is allowed to re-enter the community. Studies, including one by the Department of Justice, suggest the approach can be an effective way to decrease repeat offenses and works for perpetrators and victims.

*[As technology advances, will it continue to blur the lines between public and private? Sign up for Charlie Warzel's limited-run newsletter to explore what's at stake and what you can do about it.]*

For Ms. Blackwell, applying a similar tactic to tech platforms made sense. Current tech company enforcements, if enacted, tend to be harsh and geared toward deterrence, not treating the underlying causes of rule-breaking behavior.

Ms. Blackwell and Mr. Loewinger decided to run what they called "a highly unscientific" experiment on Reddit, a social network with tens of thousands of forum communities. Each community is policed by volunteer moderators who take down offensive posts and enforce that community's set of rules. Ms. Blackwell and Mr. Loewinger teamed up with the moderators of Reddit's r/Christianity community, which has roughly 200,000 members. It is diverse, comprising L.G.B.T.Q. Christians, fundamentalists, atheists and others with an interest in posting about the faith. Discussions get intense.

The pair selected three users who were barred for repeatedly violating rules. They created a chat room where the offender and community moderator would meet with Mr. Loewinger and Ms. Blackwell, who acted as mediators. The offenders would be confronted with past bad behavior and given the opportunity to better understand why they were barred. Upon successful completion, they'd be readmitted to the group.

**Sign up for the Kara Swisher newsletter, for Times subscribers only.** The host of the "Sway" podcast shares her insights on the changing power dynamics in tech and media. [Get it in your inbox.](#)

The results were mixed. In one case, mediation broke down, in part because of Ms. Blackwell and Mr. Loewinger's inexperience mediating and tensions between a user and a moderator that boiled over. The second case, which involved an anti-gay user who was accused of bullying an L.G.B.T.Q. user into committing suicide years ago, proved simply too toxic to

continue. The third case, involving “James,” an atheist and biblical historian who was barred for repeatedly violating r/Christianity’s rules for civil discussion, was a success.

At various points throughout the chat log of the mediation, James expressed genuine shock. “Dang this wasn’t the context that I remembered,” he types at one point, after looking at past bullying posts. “I thought someone else was the instigator and I felt ganged-up on or something. But ... looks like I was the instigator.” He apologized for lashing out, at one point suggesting “the problem is more obviously about (mis)communication and hostility that comes up in the course of these conversations.” Eventually, the moderators lifted their ban.

When I spoke to James over the phone about the process, he described his aggressive behavior as a kind of dissociation — a moment of weakness where he stopped seeing those on the other end of the thread as real people. “My frustration expressed itself as insult diarrhea with no regard to whether I was being reasonable,” he said. He noted that he’d been back in the community for two months, is more conscious of his interactions and has yet to break the rules.

James isn’t convinced the process could work for everyone. He argued that mediation was effective for his specific personality type. “It’s the element of shame,” he said. “I’m somebody who feels guilt being confronted and it allowed me to see I was the one at fault.” Ms. Blackwell and Mr. Loewinger’s mixed results suggest success is far from guaranteed. Online, mediators have to deal with pseudonymous individuals, trolls and pranksters with no desire to reform. Even those dealing in good faith might bristle at having to apologize or confront their victims. Given the nature of online harassment and bullying, the restorative justice approach is full of pitfalls. Forcing targeted minorities or vulnerable users to confront abusers, for one, could increase trauma or put undue burden on victims.

Most daunting is the issue of scale. There’s simply no way to replicate the amount of time and effort involved with Ms. Blackwell and Mr. Loewinger’s experiment across the web. “It’s like trying to moderate a wild river,” an r/Christianity moderator said in the chat logs. “It’s only getting worse, too. I can’t even begin to evaluate all of this stuff.” The ceaseless torrent of posts and comments is why tech platforms are increasingly turning to algorithms and artificial intelligence to solve the problem.

But successful moderation — the kind that not only keeps a community from collapsing under the weight of its own toxicity but also creates a healthy forum — requires a human touch. Even skilled moderators assume a huge psychological burden; many working for Facebook and YouTube are outside contractors, subjected daily to torrents of psychologically traumatizing content and almost always without proper resources. Even in small communities, keeping the peace requires a herculean effort. A recent *New Yorker* article described the job of two human moderators of a midsize tech-news message board as an act of “relentless patience and good faith.”

This reality makes Ms. Blackwell and Mr. Loewinger’s experiment equal parts compelling and dispiriting. Mr. Loewinger remains optimistic. “It’s easy to write off all people who exhibit jerk-ish behavior online as pathological trolls,” he told me. “Dislodging that assumption might hold the key to a less toxic web. The James case demonstrated to me that people are open to reflecting on what they’ve done, especially when treated with dignity.” Ms. Blackwell argued that having reformed users back in the community actually makes the forums healthier. “We will never effectively reduce online harassment unless we address the underlying motivations for participating in abusive behavior, and having reformed violators go on to model prosocial norms is an incredible bonus,” she said.

But if reform means an abundance of shame and dignity on the internet, it’s hard not to feel that all is already lost. Still, the pair’s earnestness is refreshing. And at its core there’s a lesson: If fast, scalable algorithmic solutions gave us the broken system we’ve got, it’s stripped-down patience and humanity that have the best chance of pulling us out.

*The Times is committed to publishing a diversity of letters to the editor. We’d like to hear what you think about this or any of our articles. Here are some tips. And here’s our email: [letters@nytimes.com](mailto:letters@nytimes.com).*

*Follow The New York Times Opinion section on Facebook, Twitter (@NYTopinion) and Instagram.*

Charlie Warzel, a New York Times Opinion writer at large, covers technology, media, politics and online extremism. He welcomes your tips and feedback: [charlie.warzel@nytimes.com](mailto:charlie.warzel@nytimes.com) | [@cwarzel](https://twitter.com/cwarzel)

A version of this article appears in print on , Section A, Page 22 of the New York edition with the headline: Could Restorative Justice Fix the Internet?