

 [Back to All](#)



When Kindness Kills: How algorithms accelerate savior swarms

Published on January 24, 2024

William R. Frey, J. Nathan Matias

CITATION



in

ALGORITHMS

BIPOC

BLACK TWITTER

SOCIAL MEDIA

TWITTER

In the days following the murder of George Floyd, moderators of r/BlackPeopleTwitter — a community on Reddit — went into a six-week lockdown. On a normal day, the community shares jokes, memes, and social commentary centering Blackness and Black people on topics including reproductive health, criticisms of artificial intelligence, and instances of white American hypocrisy — often through humor. But as Reddit’s algorithm recommended popular posts in the aftermath of Floyd’s murder, new users flooded into the community spewing hate and racism. When asked about the experience, one moderator, Jefferson Kelley reflected, “We were working full time to deflect users that couldn’t stand showing support for a Black man murdered by a white police officer.”

While this community lockdown enabled verified Black users to find refuge from racism surging across Reddit, carefully curating this safe space did not come without costs. Moderators were inundated with messages filled with racist slurs, and endless verification requests from professedly well-intended users. And as posts from r/BlackPeopleTwitter were elevated to Reddit’s front page, thousands of new white users filled moderator inboxes, bubbling with burgeoning racial awareness and offering unsolicited help. These moderators were experiencing a savior swarm.

Savior swarming: *An overwhelming flow of concentrated collective action, motivated by an uninformed desire to help, escalated by algorithmically-mediated communications.*

An earlier savior swarm descended on Newtown, Connecticut in December 2012, after 20 children and 6 adults were murdered in a school shooting at Sandy Hook Elementary School. As the news spread, people across the US organized on Facebook pages to offer support. Using growth-hacking techniques amplified by algorithms and broadcasters, well-meaning page creators organized a vast influx of goods that the grieving town was unprepared for. While Newtown had a population of 27,000, the city received 65,000 teddy bears, 9 tractor-trailers worth of paper snowflakes, and half a million letters — even after convincing school buses of gifts to divert elsewhere. The city eventually

incinerated warehouses of these gifts, incorporating the ash into a memorial ten years later.

Overwhelming or unwanted influxes of aspirational care aren't new. But online platforms and social algorithms are accelerating saviorism, especially white saviorism, to increasingly more overwhelming levels. This is what we call **savior swarming**.

Savior swarming is any overwhelming flow of concentrated collective action, motivated by an uninformed desire to help and escalated by algorithms. As a concept, it links individual behaviors with the systems that foster collective action — and then considers the capacity of the community at the receiving end. By showing how these pieces come together to make collective kindness unsustainable or damaging to recipients, we hope to make savior swarms clearer to recognize, helping communities protect themselves and steer swarms in beneficial ways.

Both r/BlackPeopleTwitter and the town of Newtown could handle offers of support within limits. But neither were prepared for a concentrated avalanche of thousands of people who expected to be validated by a stunned community that never asked for that kind of engagement. Savior swarming's effects on the target community aren't only the unsustainable labor, cost, and clogged logistics; savior swarms also create an unbearable and impossible-to-satisfy burden of reciprocity that weighs deeply on the mental health of the receivers. And when these swarms dissipate without taking on the deeper issues (for example, gun violence, police violence, racism), they can poison the well against future cooperation that makes solutions possible.

Saviorism

However savior swarms end, they begin with the desire to help. This saviorism is best illustrated with a picture from one of many teddy bear donation pages from the days after the Sandy Hook shooting. In the photo, a white woman smiles benevolently at the camera, half-submerged in a pile of bears that covers the couch where she is sitting. In a later post, she poses in front of a school bus with

bears up to the ceiling. These images put the helper at the center of the story, a hallmark of what scholars call saviorism.

Critiques of white saviorism can be traced throughout the history of colonialism, wherever mindsets and pathways for action elevate white people as especially capable, intelligent, and “developed.” White saviorism thrives where histories of injustice, racism, and domination have created hierarchies between the helpers and the helped in the minds of would-be saviors, what Teju Cole calls “the white savior industrial complex.”

To explain, Cole recounts the days following another savior swarm: the one in response to the Kony 2012 video launched by the non-profit Invisible Children. The Kony 2012 campaign influenced Twitter’s trending algorithm to promote an ill-conceived and ultimately disastrous humanitarian campaign. As Cole tells it, in their outsized enthusiasm for the campaign, international aid institutions and the media created yet another framework for people to use Africa as a “backdrop for white fantasies of conquest and heroism” — the perfect petri dish for saviorism. In response to the benevolent urgency encouraged by the campaign, the public quickly spread these oversimplified portrayals of Kony and Uganda, often demanding further US military involvement on African soil.

By writing their own story, modern day saviors are often on journeys to save themselves. The saviorism script offers people a way to separate themselves from responsibility for systems of harm (racism and colonialism, for example), differentiate themselves from overtly “bad” people (racists and apathetic people), cultivate self-redemptive beliefs, and gain catharsis from the guilt, shame, and fear of being seen as something other than one of the “good” people. When other people validate saviorism and join in, the result can be a swarm.

How swarms form

When Reddit users logged on in May 2020, they would have seen a feed full of news stories and perspectives on George Floyd’s murder. Reddit users, who were predominantly young white men with some college education, responded by up-voting and commenting on the articles they saw in their feeds. They

upvoted breaking news and commentary about Floyd's murder over 65,000 times. They upvoted commentary on the relationship between illegality and resistance over 82,000 times. They upvoted debates on defunding the police over 74,000 times. And they commented at volume. The result was a cycle of moral outrage and algorithm amplification that scientists describe as a feedback loop: as more people reacted, Reddit's algorithms promoted the articles to even more people.

Here's what we think likely happened next. As conversations on Reddit about George Floyd's murder spread even more widely, they were seen by an increasing proportion of people who did not have experience thinking about racial justice. With a potential audience of millions, these conversations became more valuable to anyone with an agenda, increasing the probability of both racist commentary and parachute interventions from would-be saviors.

Managing savior swarms

Savior swarms can have strong negative effects on the organizing capacity and mental health of the people who saviors are positioning themselves to help. In 2023, researchers at Pew found that many Black Americans reported negative impacts on their mental health from seeing videos of police violence played and replayed, whether on television or online. Despite this, sixty-five percent of Black Americans agreed it's a good thing that those videos were shared — sharing presumably motivated by desires to help (e.g., raise awareness) and be seen by others as doing so.

That's the dilemma for people at the receiving end of a savior swarm, who field an avalanche of sometimes painful, disruptive, uninformed content and behaviors with the hope that it achieves something positive in the long term. As r/BlackPeopleTwitter moderators on Reddit made their way through the onslaught of savior swarmers in 2020, they were able to redirect the benevolent energy of white users towards ongoing mutual aid efforts and healthy forms of collective action.

Community leaders from marginalized groups also have a long history of strategically positioning themselves in spaces where swarms of prejudiced

conversations regularly form. Michael Golebiewski and danah boyd have called these algorithmic gravity wells “data voids” and Francesca Tripodi has documented the role they play in public health and political discourse. On Reddit, one Black moderator, Jefferson Kelley, took over a data void used to mock Black parents and converted it into a space that celebrates Black families.

What the community response to savior swarms teaches us about algorithm governance

At a time when scientists and policymakers are debating the future of algorithm governance, communities like r/BlackPeopleTwitter are developing ways to describe, detect, and manage savior swarms out of necessity. As we continue to study savior swarms and community responses, we see three lessons for scientists, policymakers, and communities.

First, scientists and policymakers can benefit from listening to and recognizing the lived experience and expertise of communities most affected by algorithmic systems. When people have to live with the consequences of an algorithm, they develop creative, sophisticated ways to name, forecast, and manage the resulting dynamics. That’s true of savior swarms and so many other complex issues that scientists struggle to understand.

Second, algorithmic governance debates need to acknowledge the complex trade-offs of any policy and incorporate the wishes of communities into policy goals. Many approaches to digital protection are based on monitoring, silencing, and punishing individuals. The leaders of r/BlackPeopleTwitter saw savior swarms as a more complex collective problem. By locking down while continuing to engage with incoming requests, their solution partly protected their community from the swarm while providing pathways for well-meaning white users toward healthy collective action. It was a costly choice, but one they considered worthwhile.

Finally, scientists can advance the study of human-algorithm behavior by paying more attention to savior swarms, which have become a regularly recurring pattern in algorithmically-infused societies. By looking for more examples of these swarms, studying how they work, and working with communities to test

responses, scientists can make a difference while advancing knowledge at the same time.

William R. Frey is a PhD candidate at Columbia University's School of Social Work. His scholarship examines race and social media, drawing upon interdisciplinary understandings of race and critical framings of technology. William's research is informed by 15 years of inter-/intragroup dialogic facilitation practice focused on people's experiences with race, racism, and whiteness. He has an MSW in community organization and a BA in psychology from the University of Michigan.

Dr. J. Nathan Matias (@natematias) is a computer scientist and social scientist who organizes citizen behavioral science for a safer, fairer, more understanding Internet. A Guatemalan-American, Nathan is founder of the Citizens and Technology Lab and an assistant professor in the Cornell University Department of Communication.

Related Resources



ACADEMIC ARTICLE

April 16, 2026

Broken Connections: Fieldnotes from the Old Internet

Abstract: This essay reflects on the lived experience of early internet culture to interrogate what has been lost in the transition to today's platform-dominated online...

HISTORY OF THE INTERNET

SOCIAL MEDIA

ACADEMIC ARTICLE

April 6, 2026

Disinformation as Cultural Narrative: Conceptualizing Disinformation as Cross-Platform, Identity-Affirming, Cathartic Stories

Abstract: "Rather than framing disinformation as false facts which can be countered by true facts, we propose a model of disinformation as narrative by tracing three case studies..."

DISINFORMATION

FACEBOOK

POLITICS

...

REPORT

February 4, 2026

(404) Job Not Found: What Workforce Training Can't Fix for Black Atlantans in the Age of AI

AI LITERACY

ARTIFICIAL INTELLIGENCE

ATLANTA

...

BLOG POST

January 28, 2026

"AI Can't Do Nothing With Us!": Inside One Atlanta Workforce Training Program

ATLANTA

BIPOC

CYBERSECURITY

...

BLOG POST

January 21, 2026

How Tech Platforms Fail to Protect Societies in Conflict Regions

CONTENT MODERATION

GLOBAL MAJORITY

SOCIAL MEDIA

ARTICLE

November 14, 2025

Meta Must Rein In Scammers — Or Face Consequences

DIGITAL ADVERTISING

FACEBOOK

INSTAGRAM

...

BLOG POST

October 29, 2025

Participation and Trust: The Uses and Limits of Algorithmic Impact Assessments

ALGORITHMIC IMPACT ASSESSMENT

ALGORITHMS

IMPACT ASSESSMENT

...

ARTICLE

October 21, 2025

Spiritual Chatbots in Uncertain Times

ALGORITHMS

ARTIFICIAL INTELLIGENCE

CHATBOT

...

LOAD MORE



Explore our research programs

LEARN MORE



DATA & SOCIETY

All Data & Society, no spam.

Interested in signing up for our newsletter updates?

SUBSCRIBE TODAY



[CONTACT US](#)

[DONATE](#)

[CAREERS](#)

[ANNUAL REPORTS AND
FINANCIALS](#)

[CONTENT DISCLAIMER
AND PRIVACY POLICY](#)

[SITE BY GB](#)

© 2026 Data & Society. All rights reserved.

Unless otherwise noted, this site and its contents are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license.

