

Subverting the algorithm: Examining anti-algorithmic tactics on social media

by

Anna Woorim Chung

B.A. Media Studies
Pomona College, 2014

SUBMITTED TO THE PROGRAM IN COMPARATIVE MEDIA STUDIES/WRITING IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN COMPARATIVE MEDIA STUDIES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 2020

©2020 Anna Woorim Chung. All rights reserved.

The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part
in any medium now known or hereafter created.

Signature of Author: _____
Department of Comparative Media Studies
May 7, 2020

Certified by: _____
Ethan Zuckerman
Director, Center for Civic Media
Thesis Supervisor

Accepted by: _____
Eric Klopfer
Professor of Comparative Media Studies
Director of Graduate Studies, Comparative Media Studies

Subverting the algorithm: Examining anti-algorithmic tactics on social media

by

Anna Woorim Chung

Submitted to the Department of Comparative Media Studies on May 7, 2020 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Comparative Media Studies

Abstract

As social media users have become increasingly aware of algorithms and their potentially negative effects, I argue that some users are challenging algorithmic systems of control. This thesis examines *anti-algorithmic tactics* on online platforms -- ways in which users actively aim to subvert algorithmic systems. I investigate anti-algorithmic tactics through two case studies of social media users responding to algorithmic content moderation and content curation. The first case study investigates how Black Facebook users have used alternative spellings to avoid detection by content moderation algorithms. The second case study investigates how users of Gobo, a social media browsing tool, have used tactics to minimize the influence of content curation algorithms on their social media feeds. In these case studies, I conduct close readings of public social media posts and interviews with social media users to better understand the perceptions around algorithms and motivations for anti-algorithmic tactics. Based on insights from these case studies, I conclude this thesis by discussing design frameworks that address concerns around transparency and user agency in algorithmic systems.

Thesis Supervisor:

Ethan Zuckerman
Director, Center for Civic Media

Committee Members:

Tad Hirsch
Professor of Art + Design, Northeastern University

Nick Seaver
Professor of Anthropology, Tufts University

Table of Contents

Acknowledgements	4
Introduction	5
What do we mean by “algorithm”?	7
Defining anti-algorithmic tactics	13
Investigating anti-algorithmic tactics on social media	16
Chapter 1 - “Wypipo”, “yt”, “whyte”: Alternative spellings as tactic for subverting content moderation algorithms	19
Background	19
Methods	20
Folks theories of content moderation on Facebook	25
Alternative spellings as tactic and culture	33
Conclusion	36
Chapter 2 - “I don’t trust the algorithm”: Perceptions of content curation algorithms and alternative modes of curation	38
Background	38
Methods	44
User frustrations with algorithm-driven social feeds	46
Anti-algorithmic tactics for browsing social media content	49
Conclusion	54
Chapter 3 - Designing for user agency on algorithmic platforms	56
Seamful design: enhancing user understanding of algorithms	59
Contestability: enabling user participation in algorithmic decision-making	62
Conclusion	65
Appendix	68
References	71

Acknowledgements

This thesis reflects many of the ideas, conversations, and collaborations I shared with people during my time at MIT. I am especially grateful to have been a part of Civic Media and Comparative Media Studies.

To the Civic crew — thank you for being such a creative and supportive community. Special shoutouts go to Dennis and Rahul, for teaching me so much from our work together on Gobo; Stephanie, for being a wonderful supporter and helping me grow as a designer; and of course, Ethan, for being a fearless leader and a caring advisor. I feel endlessly grateful and honored to have been with Civic for its last semester at MIT.

To CMS '20 — thank you for being amazing classmates and friends. I couldn't imagine a better cohort to have shared my time at MIT with. Lizzi, you've brought the most positive energy and advice to my life. Ben and Sam, I enjoyed every minute of A-Side/B-Side with y'all and will sorely miss Monday nights at WMBR.

To my Boston friends — thank you for making this city feel like home. Isabel, Eyassu, Ben, Sam, Zivvy, Jaira, Tian, and the Claremont crew, you've made my two years here so special.

And last, but not least, to my family and friends outside of Boston — thank you for giving me love, support, and encouragement from day 1.

Introduction

I start off every morning by opening up Spotify on my computer and playing one of my various playlists. As I get ready for the day, the playlist eventually reaches the end, but Spotify continues to play more songs. These songs are often ones that I've heard before -- are they really playing that Mereba song *again*? I skip to the next song -- another one I've heard before. *Thanks, Spotify algorithm.*

As I pack up my bag and head over to my research lab, I put my headphones on and turn on a different playlist. While I'm waiting for the bus, I open up notifications from Instagram. After responding to a couple messages, I start browsing through my feed. Some selfies, home-cooked meals, dance videos, and ads. I stop to read one of the ads -- a late night event at the Institute of Contemporary Art -- and save the post to my account. On second thought, I also send it to two of my friends. "Plans for next weekend?" A few minutes later, my friend responds, "I'm down!" *Thanks, Instagram algorithm!*

From music recommendations to curated ads, from ranked search results to shopping recommendations, algorithms seem to be present in every corner of our online experience. A couple years ago, I was barely aware of algorithms in my life; today, I'm much more attuned to the influence of algorithms in what I watch, what I listen to, and what I share. Over the years, I've developed preferences for certain algorithms (Instagram's algorithm "gets me" more than Twitter's algorithm) and became frustrated with others (I rarely browse Facebook's algorithm-driven News Feed now).

In fact, algorithms are mediating much of the information we encounter online¹, and we're recognizing the significant social, political, and economic ramifications this can have. In

¹ Singh, "Recommendation Algorithms."

recent years, social media algorithms have been blamed for propagating misinformation², silencing marginalized voices³, and skewing political elections.⁴ Algorithms have also been blamed for creating filter bubbles⁵, the phenomenon where we only encounter a narrow range of content that reinforces our existing interests and beliefs.

On large institutional levels – in academia, government, and within tech companies themselves – we’re seeing calls to address some of the issues that social media algorithms have raised. Many of these calls recognize the tech company as both the site of the problem and the site where change needs to take place -- the tech companies should be regulated⁶ and their algorithms should be more transparent.⁷ However, as regulations on the use of algorithms are slowly being proposed from the top-down, I argue that users are already actively negotiating algorithms from the bottom up. In addition to influencers and marketers who have long engaged with approaches for “hacking” algorithms to optimize content delivery on social media platforms, everyday social media users have also shared various ways to “defy”⁸ or “avoid the algorithm.”⁹

On platforms like Facebook, Twitter, and Instagram, social media users who are frustrated with a platform’s algorithms are making changes to their settings and behaviors, as well as using third-party tools, in order to minimize the influence of algorithms on their feed. I refer to these practices as *anti-algorithmic tactics* -- ways in which users actively aim to subvert algorithmic systems. Anti-algorithmic tactics are not necessarily “successful” in subverting algorithmic systems; rather than measuring the success of a tactic, I’m interested in how an anti-algorithmic tactic is informed by a user’s understandings of and experiences with an

² Wu, “Radical Ideas.”

³ Greene, “Alex Jones.”

⁴ Hern, “Filter Bubbles.”

⁵ Pariser, *Filter Bubble*.

⁶ Applebaum, “Regulate Social Media.”

⁷ Hosanagar and Jair, “Transparency in Algorithms.”

⁸ Pinsker, “Defy the Facebook Algorithm.”

⁹ Graziano, “Avoid the Algorithm.”

algorithmic system. Through examining anti-algorithmic tactics, I argue that we can uncover nuanced understandings, frustrations, and concerns around algorithms -- and better understand the ways in which algorithmic systems are designed.

In this thesis, I examine anti-algorithmic tactics on social media and the particular understandings and experiences that drive these tactics. Through conducting close readings of social media posts and open-ended interviews with social media users, I aim to address the following key questions:

1. How do social media users understand and encounter content curation and content moderation algorithms?
2. What are the anti-algorithmic tactics that social media users are engaging with, and how are these tactics informed by user understandings of algorithms?
3. What do these tactics reveal about how algorithmic systems are designed? Through understanding the motivations and frustrations that drive anti-algorithmic tactics, how might we design algorithmic systems differently?

What do we mean by “algorithm”?

Social media platforms like Facebook are particularly important sites of algorithmic mediation, given how frequently these platforms are used and how many roles they have come to serve: as a source for news¹⁰, a medium for communication¹¹, and a marketplace for buying and selling goods.¹² As social media platforms continue to take on multiple roles, they collect increasing amounts of information on what we “comment” on, what we “like”, and who we “follow” – what media scholar Jose van Dijck would characterize as a shift from a “participatory

¹⁰ Shearer and Matsa, “News Use.”

¹¹ See <https://messenger.com>.

¹² See <https://www.facebook.com/marketplace/>.

culture to a culture of connectivity.”¹³ The sum of the actions we take on social media provides companies with massive data sets, which are then processed by *algorithms*.

In this project, it is important to establish what is meant by “algorithm” in the first place. A Merriam Webster definition tells us that an algorithm is “a step-by-step procedure for solving a problem or accomplishing some end.”¹⁴ However, the term “algorithm” is loaded with various meanings and assumptions depending on the context in which it is used. We could look to the field of computer science and human-computer interaction (HCI) for a more “formal” definition of algorithm: “a finite, discrete series of instructions that receives an input and produces an output.”¹⁵ Yet even amongst technical “experts”, the term algorithm may have a “vague, ‘non-technical’ meaning, indicating various properties of a broader ‘algorithmic system.’”¹⁶

In the context of social media and online platforms, we could look to more specific explanations of algorithmic content curation and moderation. HCI scholars Emilee Rader and Rebecca Gray define algorithmic curation as “organizing, selecting, and presenting subsets of a corpus of information for consumption” and explain that “systems like Facebook and Google (and many, many others) use algorithms as information intermediaries that determine what information should be displayed and what should be hidden.”¹⁷ Meanwhile, in regards to algorithmic content moderation, HCI scholars at the Korea Advanced Institute of Science and Technology explain, “To keep up with the immense volume of content created by users, online social platforms — like Facebook, YouTube, and Twitter — are known to train and apply machine learning algorithms by compiling large datasets of past moderation decisions on the platform.”¹⁸ Given expanding user bases and the increasing amount of time that is spent online,

¹³ Van Dijck, *The Culture of Connectivity*.

¹⁴ Merriam-Webster, “Definition of Algorithm.”

¹⁵ Hogan, “Invisible Algorithms.”

¹⁶ Seaver, “Algorithms as Culture.”

¹⁷ Rader and Gray, “Algorithmic Curation.”

¹⁸ Naher et al., “Algorithmic Understandability”, 1-2.

platforms may be employing algorithms in order to automate the process of sorting through massive amounts of content.

While these definitions and explanations of algorithms may offer a lens into how researchers are considering them, what do everyday social media users mean when they refer to an “algorithm”? How are their definitions and understandings of algorithms shaped by encounters with algorithmic systems? Rather than approaching algorithms as formal, technical objects, I aim to examine how algorithms are understood and talked about in practice -- by actual users of an algorithmic system. In part, it is often impossible to directly study algorithms on platforms like Facebook and Twitter; algorithms are seen as one of the most valuable assets of for-profit social media platforms and, as such, their inner workings are largely kept secret.¹⁹ While these “black box algorithms” are formally studied by computer science researchers through testing various inputs and examining outputs²⁰, in my research, I focus on examining how social media users *informally* develop understandings of algorithms. Specifically, I investigate how understandings are shaped by personal encounters with the platform and, importantly, how these understandings motivate changes in their individual engagement with the platform as well as participation in collective practices.

Studying algorithms “as culture”

My research draws upon the work of scholars in the social sciences who argue that it is important to examine the relationship between algorithms and people -- both creators and users of algorithmic systems. In large part, my work is inspired by anthropologist Nick Seaver’s argument in “Algorithms as culture: Some tactics for the ethnography of algorithmic systems.” In this piece, Seaver argues for an approach to examining “algorithms as culture” -- as “informed

¹⁹ Pasquale, *The Black Box Society*.

²⁰ Diakopoulos, “Accountability.”

by their empirical profusion and practical existence in the wild – always at the boundaries of diverse communities of practice.”²¹ In particular, Seaver distinguishes the approach to “algorithms as culture” from the notion of “algorithmic culture,” which “posits algorithms as a transformative force, exogenous to culture.”²² “Algorithms as culture” is also distinct from the notion of “algorithms becoming culture,” which “happens when algorithms become objects of popular debate and targets of strategic action.”²³ Seaver writes, “Algorithms are cultural not because they work on things like movies or music, or because they become objects of popular concern, but because they are composed of collective human practices. Algorithms are multiple, like culture, because they *are* culture.”²⁴

By investigating algorithms as culture -- as a collection of changing human practices rather than a static formula -- I aim to uncover the various explanations and understandings of algorithms that inform user behaviors on algorithmic systems. While Seaver’s research examines the practices of algorithmic *producers* -- engineers and research scientists who work on designing algorithms -- my own research focuses more on engaging with *users* of algorithmic systems. I focus on how users understand and relate to algorithms and argue that these relationships can provide insights into how algorithmic systems are navigated in practice, which may be different from what algorithmic producers may intend.

In “The Social Power of Algorithms,” sociologist David Beer offers another argument for examining *notions* of the algorithm, in addition to their effects. Beer writes, “We need to think not just about the impact and consequences of code, we also need to think about the powerful ways in which notions and ideas about the algorithm circulate through the social world. Within these notions of the algorithm, we are likely to find broader rationalities, knowledge-making and norms

²¹ Seaver, “Algorithms”, 5.

²² Ibid, 5.

²³ Ibid, 5.

²⁴ Ibid, 5.

– with the concept of the algorithm holding powerful and convincing sway in how things are done or how they should be done.”²⁵ While algorithms have long underpinned social media platforms, discussions around algorithms on social media have become much more widespread in recent years, particularly in the aftermath of the 2016 US presidential elections.²⁶ As algorithms are increasingly acknowledged by social media users, I believe it is important to examine how they are talked about and how “notions of the algorithm” are reshaping users’ relationships to a platform.

Uncovering algorithmic understandings through “folk theories”

Prior HCI research around “folk theories” offers a useful framework for studying user understandings of algorithms on social media platforms. HCI scholars Eslami et al. define folk theories as “non-authoritative conceptions of the world that develop among non-professionals and circulate informally.”²⁷ Folk theories are not necessarily accurate or true -- although they can be -- and may go against the intended understanding that designers aim to promote through “a system’s documentation, advertising, aesthetics, and interface.”²⁸ Rather, folk theories reflect individual understandings of a system which are derived from “first-hand experience and social interactions.”²⁹

On social media, I argue that folk theories are particularly widespread and varied due to the highly personalized, continuously changing, and relatively obscure nature of the algorithmically-driven systems for content curation and moderation. In addition, the very nature of social media, in which anyone can instantly share information, allows folk theories to quickly

²⁵ Beer, “Social Power.”

²⁶ Hern, “Filter Bubbles.”

²⁷ Eslami et al., “Folk Theories”, 2372.

²⁸ Ibid, 2372.

²⁹ Ibid, 2372.

disseminate and reach a wide range of users. Given how regularly people engage with social media and how little social media companies share about how their algorithms work, folk theories on social media algorithms can be a particularly rich area to explore algorithmic understandings.

In addition, folk theories are worth exploring not only to discover user understandings of algorithms but also to examine how these understandings inform user behavior. In their research on folk theories of Facebook's News Feed curation algorithm, Eslami et al. explain, "While the operation of these algorithms is typically opaque to users, users often develop and sometimes share theories about how these curation algorithms work in order to plan their behavior."³⁰ In other words, folk theories may be necessary in allowing a user to meaningfully navigate a platform in which the algorithmic workings are intentionally made opaque, and folk theories may provide reasonings for a user to adapt their ways of engaging with an algorithmic system. As such, HCI researchers DeVito et al. offer an expanded definition of folk theories as "intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems, which guide reactions to and behavior towards said systems."³¹ As users experience changes on social media platforms due to evolving policies, back-end systems, or interfaces, their folk theories may also evolve or expand over time.

In my research, I specifically explore folk theories of social media algorithms in order to understand how they can inform *anti-algorithmic tactics* on a platform. In particular, I am interested in identifying the experiences and sources of information that inform a user's folk theories, whether it is a specific incident or observations over time. On social media, folk theories can also be shared amongst many users and even lead to collective behavior changes

³⁰ Ibid, 2371.

³¹ DeVito et al., "Algorithms Ruin Everything," 3165.

-- which could then potentially “shape the evolution of the system as a whole.”³² Folk theories may offer a window into how social media users understand a platform’s algorithms -- and why they might try to subvert them.

Defining anti-algorithmic tactics

Through investigating folk theories and probing into user perspectives on algorithms, my goal is to understand why social media users engage with anti-algorithmic tactics. As social media users increasingly recognize the influence of algorithms on what they can see and share³³, I am interested in how negative, algorithmic encounters on a platform can lead users to adopt anti-algorithmic tactics. On social media, users may engage with anti-algorithmic tactics to avoid algorithm-driven content curation or to avoid detection by content moderation algorithms. I am interested in the ways that users engage with anti-algorithmic tactics not because they are necessarily successful in subverting an algorithmic system, but because they illuminate underlying understandings, concerns, and frustrations around algorithms.

I specifically characterize anti-algorithmic responses as *tactics* to underscore the power relations between users and an algorithm-driven platform. In *The Practice of Everyday Life*, Michel de Certeau lays out a distinction between strategies and tactics. Using this distinction, I argue that platforms employ algorithms as a *strategy* while platform users employ anti-algorithmic *tactics* to subvert these algorithms. De Certeau calls a strategy “the calculation (or manipulation) of power relationships that becomes possible as soon as a subject with will and power (a business, an army, a city, a scientific institution) can be isolated. It postulates a *place* that can be delimited as its *own* and serve as the base from which relations with an

³² Eslami et al., “Folk theories”, 2371.

³³ Smith, “Public Attitudes.”

exteriority composed of targets or threats (customers or competitors, etc.) can be managed.”³⁴

On the other hand, a tactic is “determined by the *absence of power* just as a strategy is organized by the postulation of power.”³⁵ Tactics may emerge in direct response to strategies and, as a result, can be more adaptable to different contexts. As de Certeau writes, a tactic “must vigilantly make use of the cracks that particular conjunctions open in the surveillance of the proprietary powers.”³⁶

For example, a platform like Facebook may be seen as the powerful business which manages its users through content moderation algorithms, which can efficiently detect posts and users that violate the platform’s community guidelines. Content moderation algorithms may be seen as a strategy of surveillance, which enables Facebook to monitor and manage its users at scale. However, these algorithms might fail to identify all potential violations of these guidelines, given the sheer heterogeneity of users as well as the ability of these users to adapt to platform strategies. A strategy assumes that subjects can be managed within “a place that can be delimited as its *own*” but does not necessarily account for how subjects can respond and adapt with their own tactics. In my research, I use de Certeau’s concept to explore how anti-algorithmic tactics are informed by understandings of an algorithmic system and its “cracks” -- ways in which it might be possible to avoid algorithmic detection or undermine algorithmic influence.

Past work on “algorithmic resistance”

My approach to studying anti-algorithmic tactics draws upon prior work examining algorithmic resistance on online platforms. Media scholars Julia Velkova and Anne Kaun’s

³⁴ de Certeau, *Everyday Life*, 36.

³⁵ *Ibid*, 38.

³⁶ *Ibid*, 37.

piece, “Algorithmic resistance: media practices and the politics of repair”, has particularly served as a source of inspiration for closely examining user responses to algorithms and possibilities for user agency on algorithmic systems. Velkova and Kaun write, “As algorithms assume a dominant role in the mediation of power, it becomes increasingly important to consider to what extent and in what ways their power can be resisted.”³⁷ In their work, Velkova and Kaun recognize the increasingly active roles that users play in “shaping the workings of algorithms” and their strategic interventions in “the algorithmic politics of attention.”³⁸

Velkova and Kaun are particularly concerned with “media practices of repair”, or “tactics to correct existing shortcomings within algorithmic culture rather than by producing alternative pathways.”³⁹ They illustrate this idea through a case study on Swedish artist, Johanna Burai, who carried out a media campaign to intervene in Google Image’s algorithm-driven search results. While searching for images of hands, Burai found that Google’s results were racially biased, as all of the top results showed only images of white hands. Driven to action, Burai aimed to reconfigure Google’s algorithm-driven results through creating a website to offer images of non-white hands and carrying out a media campaign to boost these images in Google’s search ranking. Through this case study, Velkova and Kaun examine the way a user is motivated and driven to “repair” an algorithmic system through generating “alternative outputs.”

⁴⁰ Rather than rejecting an algorithmic system entirely or creating a new system, “media practices of repair” work within what Velkova and Kaun call the “algorithmic logics” of a system.

41

³⁷ Velkova and Kaun, “Algorithmic Resistance”, 2.

³⁸ Ibid, 4.

³⁹ Ibid, 1-2.

⁴⁰ Velkova and Kaun, “Algorithmic Resistance,” 13.

⁴¹ Velkova and Kaun, “Algorithmic Resistance,” 3.

Velkova and Kaun's "media practices of repair" offers a compelling way to examine how users can actively reshape and influence the outputs of an algorithmic system. However, in contrast to their exploration of "practices of repair", I am interested in *tactics of subversion* -- that is, the ways in which users, rather than reshaping algorithmic outputs, aim to avoid or undermine the influence of algorithms on a platform. While "practices of repair" entail more organized, coordinated responses to correct biases of an algorithmic system, I argue that anti-algorithmic tactics can take place more organically, motivated by critiques of an algorithmic system as well as personal preferences -- simply not liking certain algorithmic outputs. Anti-algorithmic tactics might include both practices of both repair and subversion, but in my work, I focus on tactics of subversion in order to probe further into the ways that users are actively resisting algorithmic influence online.

Investigating anti-algorithmic tactics on social media

My research explores anti-algorithmic tactics through two case studies, each exploring a group of social media users motivated by different experiences with algorithms. I selected these particular case studies to examine different types of *algorithms* that social media users encounter, different *motivations* for challenging the algorithms, and different types of *tactics*. I am interested in both identifying these differences as well as connecting these case studies through an understanding of how anti-algorithmic tactics reflect broader concerns around the power that social media platforms hold through invisible algorithms -- and the efforts to reclaim agency on these platforms.

In the first chapter, I present a case study on how a group of Facebook users have used alternative spellings to subvert content moderation algorithms. Alternative spellings for white people, such as "wypipo", "yt", and "whyte", are commonly used by Black social media users to

express humorous observations of white culture and critiques of white supremacy. However, for a subset of these users, these alternative spellings are also explicitly used to avoid account bans from content moderation algorithms. This motivation for alternative spellings emerged after several black Facebook users had started sharing experiences of being reported and banned from the platform due to posting content with critical mentions of “white people.” In many cases, these users had shared critical views on racism and white supremacy -- views which are perhaps divisive but not prohibited from being shared under Facebook’s content moderation guidelines.

Several users believed that Facebook’s content moderation algorithms were automatically flagging posts that mentioned “white people” and, in response, intentionally began using alternative spellings of “white people” to avoid further censorship. In order to better understand what motivated alternative spellings as an anti-algorithmic tactic, I collected public Facebook and Twitter posts that mention alternative spellings as a response to Facebook’s content moderation algorithms. I then conducted a close reading of these posts to unpack several “folk theories” around Facebook’s content moderation algorithms.

In the second chapter, I present a case study that examines how users of Gobo, a social media browsing tool, have aimed to subvert content curation algorithms through using platform features and third-party tools. Started in 2017, Gobo is a tool that allows users to interrogate and “control the algorithms” that curate content on their social media feeds.⁴² A year later, I joined Gobo as a designer and researcher to learn more about Gobo users’ frustrations with browsing social media feeds and to understand what motivated them to use Gobo. Along with Gobo’s lead developer, I conducted interviews with 10 Gobo users and probed into their relationships with algorithm-driven social media feeds. In particular, I aimed to understand how

⁴² See <https://gobo.social/>.

frustrations and folk theories around algorithm-driven feeds led to particular tactics to avoid or minimize algorithmic content curation.

In the third chapter, I explore tools and design frameworks to consider how algorithmic systems could be designed to give users greater agency over algorithmic decision making. As users are starting to locate platform power in algorithms, I argue that the ways in which they engage with platforms are shaped by their understanding of how algorithms operate. Examining anti-algorithmic tactics can reveal the ways in which algorithm-driven platforms inhibit transparency and user agency and lead to the marginalization of certain users. Ultimately, I argue that anti-algorithmic tactics can illuminate relationships that users have with algorithmic systems and inform thinking around how algorithmic systems could be designed differently.

Chapter 1

“Wypipo”, “yt”, “whyte”: Alternative spellings as tactic for subverting content moderation algorithms

Background

In 2017, several Black Facebook users began sharing their experiences about being banned from the platform.⁴³ Some users noted that the posts that led to their bans contained critical observations about race and mentioned the keywords, “white people.”⁴⁴ Facebook itself offered little explanation for how and why posts were being flagged, sometimes simply stating that a post violates “community standards.” Some users speculated that white supremacists had actively searched for posts that mentioned “white people” in a critical light, and reported it to the platform⁴⁵ -- a strategy that has frequently been used by trolls on the web.⁴⁶ But other users suspected that these posts were automatically being flagged by the platform with the help of algorithms -- particularly algorithms which detected posts for keywords like “white people.”⁴⁷

Given how little Facebook reveals about how its content moderation algorithms actually work, users often rely on “folk theories” -- understandings of how a platform’s algorithms work based on personal experiences, social interactions, and information encountered online. In this case, folk theories suggested that Facebook uses algorithms to flag posts that potentially violate the platform’s content moderation guidelines through the detection of certain keywords. These

⁴³ Source:

https://www.facebook.com/permalink.php?story_fbid=1755546164742468&id=100008613040087

⁴⁴ Source: <https://www.facebook.com/lf.melton/posts/316402978804966>

⁴⁵ Source: <https://www.facebook.com/stjblondell/posts/10102563264788763>

⁴⁶ Notopoulos, “Trolls Locked My Twitter Account.”

⁴⁷ Source: <https://www.facebook.com/lf.melton/posts/316402978804966>

folk theories can be understood as the reasons for why some social media users had begun using alternative spellings of “white people.” In this way, “wypipo”, “yt”, “whyte”, and “hwhite” can be seen as a way for users to avoid detection by the algorithms -- as well as a way to express critiques and humorous observations about white culture.

To better understand why alternative spellings have been used as an anti-algorithmic tactic, I identify several folk theories that have been shared by Facebook users and unpack their context and meanings. In particular, I focus on the following questions:

1. Why have alternative spellings been used as an anti-algorithmic tactic for subverting content moderation algorithms? How have specific folk theories have informed this tactic?
2. How do the experiences of Black Facebook users illuminate concerns around marginalization by algorithms on social media?

Methods

Through a survey of public posts from Facebook and Twitter, I particularly explore how alternative spellings have been used and identify several folk theories of Facebook’s content moderation algorithms. I collected Facebook and Twitter posts through conducting keyword searches for alternatives spellings of “white” and “white people”, including “wypipo”, “yt”, “hwhite”, “whyte”, and “hwhite.” In addition, I conducted keywords searches of these spellings along with “algorithm” to understand how alternative spellings were specifically talked about in relation to platform algorithms. After collecting the posts, I conducted close readings of the posts to further unpack their contexts and meanings. For posts that included media attachments, I also conducted close readings of photos and articles, and for posts that had engagement from other users, I further investigated a post’s comment threads.

Alternative spellings of “white people”: origins and uses

On social media, “wypipo” and “yt people” have often appeared in place of the phrase “white people.” Across Twitter alone, these spellings have been used thousands of times,⁴⁸ with the earliest known usage⁴⁹ being from Twitter user @DVSblast on October 18, 2012 (Image 1).⁵⁰ Since then, wypipo, along with other alternative spellings for white people, has been adopted widely online, particularly in Black communities in which alternative spellings, or what have also been called “non-standard spellings”, are a common linguistic practice.⁵¹



Image 1: The earliest known mention of “wypipo” on Twitter, shared by user @DVSblast in 2012.⁵²

One way that alternative spellings of white people have frequently been used is to share humorous observations of white culture. These observations often poke fun at stereotypes of white culture from a position of a person of color. For example, Twitter user @sleepyzee responded to a Tweet containing an image of unseasoned chicken and vegetables, writing “this plate so unseasoned.. only wypipo should know what it taste like” (Image 2).⁵³ This tweet addresses a stereotype about food from white cultures lacking seasoning and became one of

⁴⁸ Based on a query of all public tweets mentioning “wypipo” between 2012 and 2020.

⁴⁹ Know Your Meme, “Wypipo.”

⁵⁰ Source: <https://twitter.com/DVSblast/status/259003119072071680>

⁵¹ Dictionary.com, “Yt Folx.”

⁵² Source: <https://twitter.com/DVSblast/status/259003119072071680>

⁵³ Source: <https://twitter.com/StarDustFifi/status/1019662791395901440>

the most popular public Tweets mentioning “wypipo,” with over 3,800 likes. Another popular Tweet from @_benjvmins_ responded to a selfie posted by country singer Billy Ray Cyrus, who has golden brown hair and a dark beard. The Tweet reads, “this what yt people think jesus looks like,” and was liked over 700 times (Image 3).⁵⁴ In this Tweet, the alternative spelling “yt people” is used to address a commonly-held belief in white religious communities that Jesus was white. Both of these tweets reflect the use of alternative spellings in a humorous manner, particularly through shared cultural identification.



Image 2

⁵⁴ Source: https://twitter.com/benjvmins_/status/913512228128677888



Image 3

However, alternative spellings of white people have also been used in a more serious context to express critiques of white supremacy. These critiques sometimes follow a similar format of beginning with “dear wypipo” or “dear yt people” to confront a particular facet of white supremacy. For example, Twitter user @iHartEricka shared, “Dear YT people, instead of sharing white washed MLK quotes tomorrow, give some \$ to a black person. Reparations or bust” (Image 4).⁵⁵ This Tweet critiques the tradition of celebrating Black revolutionary leaders in America through a holiday like Martin Luther King Jr. Day, even as Black Americans continue to

⁵⁵ Source: <https://twitter.com/iHartEricka/status/1087078792005324800>

be financially marginalized. Another Twitter user @austinchanning confronts white supremacy directly in a Tweet that mentions wypipo: “Wypipo PSA: Unlearn white supremacy and purposefully dismantle it in your kids lives, or else displaying white supremacy will be a rite of passage in your child’s development as it has been for generations. Period” (Image 5).⁵⁶ Other Tweets that contain alternative spellings address issues such as cultural appropriation⁵⁷, hair policing⁵⁸, colonization⁵⁹, and gentrification⁶⁰. Throughout these tweets, “wypipo” and “yt people” are used in critical modes to talk about ways that Black people have been marginalized in America.



Image 4



Image 5

In both types of use, humorous and critical, alternative spellings of white people have come to signal a way of talking about race as a Black person online -- but another look into

⁵⁶ Source: <https://twitter.com/austinchanning/status/1086688549250969601>

⁵⁷ Source: <https://twitter.com/eveewing/status/1086824410231853062>

⁵⁸ Source: <https://twitter.com/itstyramonet/status/1105917502587105282>

⁵⁹ Source: <https://twitter.com/AdrianCJax/status/945839958077689867>

⁶⁰ Source: https://twitter.com/awkward_duck/status/831219019181076480

these spellings reveals a different, or perhaps additional, motivation for using these spellings. On Urban Dictionary, a crowdsourced online dictionary, the top definition for *wypipo* was published on January 29, 2016 and is described as “Twitter slang or dialect that with read aloud sounds like ‘white people’ which is its actual meaning.”⁶¹ However, the second definition, published around two years later on December 3, 2017, offers an explanation for why “*wypipo*” has been undertaken as a term: “to avoid detection from social media policies that unjustly ban, delete, block and suspend melanated people for criticizing the actions, the people or the behaviors of the dominate society.”⁶² Other definitions for alternative spellings such as “*whyte*” offer similar explanations: “used by writers of color when discussing racism on social media so their posts won’t be censored and their accounts suspended due to computer algorithms.”⁶³

These explanations do not necessarily explain why alternative spellings for white people have been popularized on social media, but they do begin to reveal a tension in the ways that Black people have encountered social media platforms. I argue that this tension is critical to understanding why users might aim to subvert a platform and why, in this case, the mode of subversion has been aimed as an *algorithmic* one. While alternative spellings may have been adopted for various different reasons, the specific motivation to avoid detection and censorship may uniquely point to the ways in which Black people have understood and contended with marginalization by social media platforms.

Folk theories of content moderation on Facebook

“I got zucked for 30 days because I called some awful, rude women ‘white trash’ - yt or *wypipo* is necessary to avoid the algorithm.”⁶⁴ In this tweet, user @Sounds_Like_Kat shared

⁶¹ yomamahakim, “Wypipo.”

⁶² Ibid.

⁶³ Annikeeper, “Whyte.”

⁶⁴ Source: https://twitter.com/Sounds_Like_Kat/status/1136321010871914498

that they had been “zucked”⁶⁵ -- a term derived from Facebook CEO Mark Zuckerberg’s name to describe having content taken down or being banned from Facebook.⁶⁶ Specifically, @Sounds_Like_Kat believed that they had been unfairly banned for using the term “white trash”, and they shared their experience in response to another Twitter user’s observation that “‘white’ is now a racist slur.”⁶⁷ @Sounds_Like_Kat then suggested that using alternative spellings such as “yt or wypipo is necessary to avoid the algorithm” and, subsequently, to avoid account bans.

Several other tweets^{68,69,70} that mention alternative spellings as a way to “avoid the algorithm” refer to a USA Today article published on April 24, 2019, which describes how Black activists were getting “zucked” for posts that mention “white people” in a critical context. The article explains, “So to avoid being flagged, they use digital slang such as ‘wypipo,’ emojis or hashtags to elude Facebook’s computer algorithms and content moderators.”⁷¹ The “algorithm” referred to in @Sounds_Like_Kat, then, can be inferred as Facebook’s content moderation algorithms, and the alternative spellings, or “digital slang,” can be seen as a response to these algorithms. However, it is unclear how exactly Facebook users have understood how these algorithms work and how they might be seen as racially biased.

The following analysis aims to further unpack how alternative spellings have been motivated by understandings of Facebook’s content moderation algorithms through analyzing three user folk theories about how the algorithms work. Folk theories are described by Eslami et al. as those “non-authoritative conceptions of the world that develop among non-professionals

⁶⁵ Also spelled “zucked.”

⁶⁶ Hathaway, “Mark Zuckerberg.”

⁶⁷ Source: <https://twitter.com/PollyTickal/status/1136314813691260929>

⁶⁸ Source: <https://twitter.com/triketora/status/1121458854607425536>

⁶⁹ Source: <https://twitter.com/harlanyu/status/1121444941085982720>

⁷⁰ Source: <https://twitter.com/bgyrl4life/status/1185178656601268226>

⁷¹ Guynn, “Facebook While Black.”

and circulate informally...via first-hand experience and social interactions.”⁷² As such, I aimed to uncover folk theories that show how Facebook users have understood the platform’s content moderation algorithms and their racialized nature through investigating shared experiences and social interactions on the platform.

The folk theories I identified were based on three main sources of knowledge: personal experiences with account bans, experiences with account bans shared by other users, and news articles. None of the theories were explained through Facebook’s official guidelines on content moderation, perhaps unsurprisingly, given the vague and non-transparent nature of their content moderation policy. While folk theories do not necessarily reveal what Facebook’s exact content moderation guidelines are, they offer insights into the ways that user understandings of content moderation have informed responses to the platform.

Theory 1: Facebook detects posts with keywords using site-wide algorithms

In explanations for how Facebook’s content moderation works, some users have suggested that Facebook uses an algorithm that automatically detects and flags posts with certain keywords. For Facebook user Andy Wilson, this theory of algorithmic detection has been informed through having witnessed other users’ posts in a private group being “taken down” for being “critical of white men” (Image 6).⁷³ In the comment thread of the post, Wilson explains how it is unlikely that the posts were reported by people within the group, given that it is “private” and “well protected.” Removing the possibility of human-reported content moderation, Wilson hypothesizes that the posts were “automatically” detected by Facebook’s algorithms, which is further affirmed by another user who commented on Wilson’s post, “So weird it’s happening in private groups. That points to site wide algorithm and not just reporting” (Image 7).

⁷² Eslami et al., “Folk Theories”, 2372.

⁷³ Source: <https://www.facebook.com/whillice/posts/10102353143413640>

For Wilson, the theory of a site-wide algorithm has motivated the use of alternative spellings like “wypipo” and “yt men” even within private groups. The alternative spellings are viewed as a way to “talk in code” about “white people” and “white men” in posts that might otherwise be removed by Facebook’s algorithms. Through this theory, the goal of the “algorithm” that social media users are trying to avoid is made more clear: to detect keywords which signal content violating Facebook’s content moderation guidelines. Thus, alternative spellings are seen as helping Facebook users to avoid keyword detection by the algorithm while continuing to communicate freely on the platform. Both the goal of the algorithm as well as its scope -- in this theory, the algorithm is used “site wide” -- is assumed based on Wilson’s observations of the kinds of posts that have been taken down in his Facebook community.



Image 6



Image 7

Theory 2: Facebook's content moderation algorithms specifically protect white people

In response to the previous theory, Facebook user LF Melton suggests that Facebook's algorithm is programmed to detect keywords to specifically protect white people (Image 8).⁷⁴ Melton specifically shared screenshots highlighting sections from a ProPublica investigative report (Image 9), which shed light on the guidelines Facebook was using to "distinguish between hate speech and legitimate political expression."⁷⁵ The article particularly revealed how Facebook's "protected categories" rule bans hate speech directed against populations based on "race, sex, gender identity, religious affiliation, national origin, ethnicity, sexual orientation and serious disability/disease (i.e. white men) but not on subsets of these populations that are not covered under protected categories (i.e. Black children and female drivers)."⁷⁶ For Melton, the protected categories rule provides a reason as to why bans could take place amongst users who engaged in commentaries and critiques of white people.

Given Facebook's protected categories, Melton believes that posts specifically containing the keyword "white" are susceptible to being identified by "Fb's algorithm" and being

⁷⁴ Source: <https://www.facebook.com/lf.melton/posts/316402978804966>

⁷⁵ Angwin and Grassegger, "Facebook's Secret Censorship."

⁷⁶ Ibid.

banned. Alternative spellings like “yt”, “wh!te”, “wyt”, and “wypipo” may be ways to avoid using the word “white”, which can be more easily seen as violating Facebook’s hate speech guidelines. Unlike Wilson’s previous theory, which was informed by their own experiences on Facebook, Melton’s theory is further informed by insights from an external source. In this way, folk theories about Facebook’s content moderation algorithms can be formed by both personal encounters on the platform as well as information gleaned through articles and reports.

Melton’s theory offers not only an understanding of how Facebook’s content moderation algorithm works but also a reason for why it works the way it does. Though Melton’s interpretation of Facebook’s protected categories rule may not accurately reflect how it actually functions, it has strongly informed their own understanding of Facebook’s content moderation policies. Based on their own observations and readings of media reports, Melton believes that Facebook’s content moderation algorithm is racially biased in specifically protecting white users. As a result, Melton sees alternative spellings of “white people” as being necessary to avoid detection by this algorithm.



Image 8

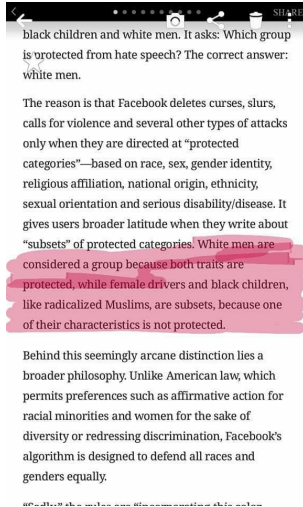


Image 9

Theory 3: Posts containing certain keywords can result in selective account bans for Black users

In addition to algorithms that lead to post removals based on the detection of keywords, another folk theory points to a more extreme action -- account bans -- which can result from criticizing white people on the platform. On April 30, 2019, Facebook user StJohnn Blondell shared their own experience of being banned from Facebook after posting a quote from prominent Black activist, Shaun King (Image 10).⁷⁷ Similarly to the previous theory, Blondell theorizes that Facebook “monitor[s] people’s posts based on keywords” -- in the case of their post, Blondell speculates that the terms, “White”, “monster”, “#blacklivesmatter”, and “#BLM” are among the keywords that may be specifically monitored. Blondell believes that the detection of one of these keywords directly led to their three day ban from the platform.

Blondell’s post was shared with an article from the Daily Dot titled, “Facebook is punishing Black people for talking about racism”⁷⁸, which refers to the earlier USA Today article

⁷⁷ Source: <https://www.facebook.com/stjblondell/posts/10102563264788763>

⁷⁸ Sadeque, “Facebook Is Punishing Black People.”

on Black activists getting “zucked.” For Blondell, who identifies as Black, the article gives further reason to believe that Black users are specifically targeted with account bans even when posting the same content as white users. Blondell shares how this theory was personally confirmed after observing that their wife, who is white, shared the same content without being banned. In Blondell’s theory of content moderation on Facebook, keyword detection plays a role in identifying posts containing a “bannable” offense, but only posts shared by Black users lead to the consequence of a banned account.

StJohnn Blondell
April 30 · 🌐

Dear #Facebook,

I have just recently finished my 3 day ban from posting (and this is my first posting), because something I posted was deemed “hate speech.” As it is favorably said by wypipo in America, “I can’t be racist, I’m friends with (insert person of color’s race).” In this case, I am friends and married to a white person.

But lets be serious here, because it’s obvious that you monitor people’s posts based on keywords. You banned me for a post that was a quote by #ShaunKing. I am figuring the use of the term “White” and the term “Monster,” tightened someone’s sphincter, or possibly it was #blacklivesmatter or #BLM that caused you to be #triggered. What is interesting is that hours after you banned me, even after I asked for a review (so the idea of an algorithm being the case is bullshit), a 19yo White male walked in to a synagogue in California and opened fired on men, women, children and the elderly after posting a manifesto that read much like the shooter in New Zealand, Pittsburgh and Norway (all White men). I think it’s safe to say that this man is a monster and he happened to also be White. He also was arrested without incident and is still alive. Why? Because American police treat such individuals with kid gloves.... which is what that quote that so offended you spoke about and was proven correct only hours after someone at Facebook was offended.

Now what is interesting is that after I was banned, a friend of mine sent me the article below which shows that Facebook has been attempting to silence People of Color that speak out against #WhiteSupremacy. It appears that as much as Facebook talks publicly about banning Hate Speech and White Supremacy, it seems that you are actually facilitating it. I am guessing the person that did not like my post probably owns a MAGA hat, and says racist things on 4Chan, 8Chan, reddit, with their friends and believes in reverse racism. Perhaps hiring more People of Color, they would be able to understand context, because the person that blocked me is ignorant at best, racist at worst. The article also points out that White people that speak out against White Supremacy don’t get banned by Facebook...I know this because my wife posted basically the same thing after you banned me and nothing happened to her. Maybe it’s just the thought of the stereotypical “uppity negro” that so offends your sensibilities. How dare us Black people speak out against injustices....after all you “freed” us, and allow us to play sports, sing, dance and make money; we should be happy with our place in life. Is that correct? Of course it is, because I have read that on multiple Facebook posts from racists that are allowed to freely claim these things all over Facebook with no consequence.

What you have proven by your despicable actions is that more of our Caucasian brothers and sisters must speak out against homophobia, transphobia, anti-Semitism, racism, etc...because you don’t ban them. We also know these things must happen, because racism in this country definitively was not created by People of Color, so it won’t be stopped by People of Color either.

You can silence People of Color speaking out against White Supremacy, because you wish to extend it, but we will continue to speak out in the streets, on the news and on social media sites like yours and Instagram. Maybe when you say Facebook attempts to “Do Better,” you should start by not keeping white supremacists working at Facebook or rather not hire them to begin with.

#hateistaught #diversity #representationmatters #equality #downwithWhiteSupremacy #freespeech

DAILYDOT.COM

About this website

Facebook is punishing Black people for talking about racism
Users say they were suspended for writing 'white people.'

👍❤️😡 27 10 Comments 20 Shares

👍 Like ➦ Share

Image 10

While these folk theories do not provide accurate or complete understandings of how Facebook's content moderation works, they do provide insights into why alternative spellings have emerged as a particular strategy to avoid account bans. The first theory suggests that Facebook uses a sitewide algorithm to detect posts with certain keywords, and the second theory more specifically suggests how this keyword detection is aimed at protecting white people. The third theory also affirms the relevance of keywords detection but argues that algorithms alone are not responsible for account bans; rather, account bans are further determined by human moderators and intentionally target Black users. Through these theories, alternative spellings may be seen as a particular response to the algorithmic, racialized nature of Facebook's content moderation process.

Alternative spellings as tactic and culture

Folk theories reveal how some Facebook users have understood the role and nature of content moderation algorithms -- and specifically the ways in which they are anti-Black. "Dark sousveillance", a concept introduced by surveillance and Black studies scholar Simone Browne, may offer one way to understand how alternative spellings serve as a tactic to avoid algorithms and, furthermore, to avoid surveillance on Facebook. Dark sousveillance is described as "a way to situate the tactics employed to render one's self out of sight, and strategies used in the flight to freedom from slavery as necessarily ones of oversight."⁷⁹ Browne provides an understanding of dark sousveillance through examples of strategies used by runaway slaves such as Sall, who was able to evade surveillance through feigning whiteness. "Sall's ability to

⁷⁹ Browne, *Dark Matters*, 21.

evade surveillance through makeup, wicked tricks, and hiding in plain sight exposes the one-drop rule as a social construction that, for some, could be subverted by performing whiteness.”⁸⁰ Dark sousveillance can provide a framework for understanding how social media users respond to content moderation algorithms, which may claim to mitigate false or harmful content but then serve to police the language and content of Black users.

However, alternative spellings may be used not only because they could subvert Facebook’s algorithms but also because they follow a history of alternative spellings being used in Black communities online.⁸¹ Alternative spellings have been seen as a characteristic of AAVE -- or African American Vernacular English -- on social media. African American Vernacular English originated in the 17th century in the American South, when Black slaves and indentured servants developed a dialect which combined British English with African and Carribean creole.⁸² While AAVE emerged in part out of a lack of educational access, it also functioned as a mode of resistance — “a covert, often defiant response to the surveillance state of slavery.”⁸³

Media studies scholar Sarah Florini argues that on Twitter, AAVE is often used in the practice of “signifyin’,” “which deploys figurative language, indirectness, doubleness, and wordplay as a means of conveying multiple layers of meaning.”⁸⁴ While it is important to acknowledge that there is no single Black cultural identity, the practice of signifyin’ through AAVE can be one important way in which Black people can perform their identity. Florini further builds upon the work of Lisa Nakamura, a leading scholar on race in digital media, arguing that this performance is “an important mode of resistance to marginalization and erasure,”⁸⁵ particularly in an online space where race could otherwise be hidden. Seen through this lens,

⁸⁰ Browne, *Dark Matters*, 54.

⁸¹ Dictionary.com, “Yt Folx.”

⁸² Winford, “African American Vernacular English.”

⁸³ Cunningham, “Black English.”

⁸⁴ Florini, “Signifyin’”, 223.

⁸⁵ Florini, “Signifyin’”, 225.

alternative spellings of “white people” may serve not only as an anti-algorithmic tactic but also as a practice of signifyin’ Black identity online -- particularly under a context in which Black users may feel targeted and marginalized by a platform.

As such, alternative spellings may serve to reduce visibility by algorithms -- and other users and human moderators -- but increase visibility amongst Black users. Alternative spellings may then be understood through the lens of “surreptitious communication design”, a concept introduced by design scholar Tad Hirsch. Surreptitious communication design “seeks to create messages that are meaningful for intended recipients, but illegible and/or inaccessible for adversaries.”⁸⁶ In this case, alternative spellings may serve not only as a tactic to subvert algorithms but also as a tactic for surreptitious communication -- to create “opportunities for vulnerable people to communicate.”⁸⁷ The use of spellings like “wypipo” and “yt people” may be recognizable but less meaningful for those outside of Black communities, which have developed collective understandings of what these spellings mean and collective practices around how they are used.

It is important to acknowledge that alternative spellings may not be an effective or sustainable anti-algorithmic tactic given that platforms could, in theory, easily modify their algorithms to detect alternative spellings of “white people.” However, while alternative spellings may not successfully subvert Facebook’s algorithms, they just as importantly *signal* subversion of the “algorithms”, which could represent Facebook’s content moderation practices more broadly. On Facebook, folk theories revealed how users viewed the platform’s content moderation guidelines as being biased against Black users while protecting white users. In addition to using alternative spellings to avoid content moderation algorithms, Black social media users have also used alternative spellings to share critical content around white culture.

⁸⁶ Hirsch, “Surreptitious Communication Design”, 65.

⁸⁷ *Ibid*, 76.

While alternative spellings of white people may serve as an anti-algorithmic tactic but, perhaps more importantly, they also serve as a way to call out anti-Blackness on Facebook.

Conclusion

The folk theories described in this case study illuminate how an anti-algorithmic tactic is shaped by an understanding of how the algorithm operates. These theories show how Facebook's content moderation algorithm has been understood as detecting posts with critical mentions of "white people." Subsequently, this algorithm has been viewed as directly leading to posts being flagged and accounts being banned by the platform. While one theory questions whether the algorithm alone is responsible for account bans, it continues to uphold the idea that certain keywords can "trigger" the platform's algorithms and make subsequent action more likely. The theory of a keyword detection algorithm can be seen as informing alternative spellings as an anti-algorithmic tactic.

Anti-algorithmic tactics may also be seen by users as a way to combat marginalization by a platform. Facebook's content moderation algorithms have been viewed as specifically marginalizing Black users for talking about race, and as a result, alternative spellings have been taken up by Black users who have already been banned before or recognize the threat of being banned by the platform. In this way, alternative spellings may be specifically motivated by the marginalization -- or threat of marginalization -- by a platform via its algorithms. Alternative spellings are not a new phenomena, and in fact, they have been used to thwart detection and censorship in the past. For example, leetspeak has been used in hacker circles since the 1980s to prevent outsiders from finding information.⁸⁸ More recently, the Mandarin word for "river crab" has been used as a code word by Chinese netizens to talk about government censorship.⁸⁹

⁸⁸ Alpine Security, "Leetspeak."

⁸⁹ Qiang, "Grass-Mud Horse."

The tactic of using alternative spellings also has a specific history rooted in the language and experiences of Black American communities. Alternative spellings are not an anomaly in the everyday linguistic practices of Black people online, and alternative spellings of white people have been used more commonly as a way of talking about race through critique and humor than as an explicit tactic of algorithmic subversion. Understanding these spellings as an anti-algorithmic tactic, though, provides important insights into the ways that people on social media are understanding marginalization as happening through algorithmic rather than human-mediated means. On social media platforms, the experiences of marginalization shared by Black users reflect wider systems of anti-Blackness in America, which are increasingly made more efficient by algorithms.

However, as users of algorithmically-enabled platforms increasingly recognize potential biases encoded in them, anti-algorithmic tactics can offer a way to navigate and express resistance on these platforms. Particularly on social media platforms, users may find ways to communicate about and shape collective practices which address concerns around algorithms. Through exploring the origins, motivations, and uses of alternative spellings, I show how folk theories of algorithms and existing cultural practices can inform a specific anti-algorithmic tactic and reveal how users negotiate algorithmic systems -- particularly systems in which they feel marginalized.

Chapter 2

“Avoiding the algorithm”: Tactics for alternative content browsing amongst Gobo users

Background

On September 5, 2006, Facebook introduced what some might consider the defining feature of their platform: the News Feed⁹⁰. The News Feed began to show every interaction our friends were having on Facebook, whether it was liked pages, relationship updates, or profile picture changes. From visual design changes to added features, such as the now iconic “like” button, the News Feed continues to evolve. But I argue that the most significant change to the News Feed — and to modern social media more broadly — took place on September 20, 2011 with the introduction of the algorithmically curated News Feed. In a product update written by Mark Tonkelowitz, an Engineering Manager at Facebook at the time, the News Feed was described as a “personal newspaper”:

“When you pick up a newspaper after not reading it for a week, the front page quickly clues you into the most interesting stories. In the past, News Feed hasn’t worked like that. Updates slide down in chronological order so it’s tough to zero in on what matters most. Now, News Feed will act more like your own personal newspaper. You won’t have to worry about missing important stuff. All your news will be in a single stream with the most interesting stories featured at the top.”⁹¹

⁹⁰ Sanghvi, “Facebook Gets a Facelift.”

⁹¹ Source:

<https://web.archive.org/web/20110925211838/http://blog.facebook.com/blog.php?post=10150286921207131>

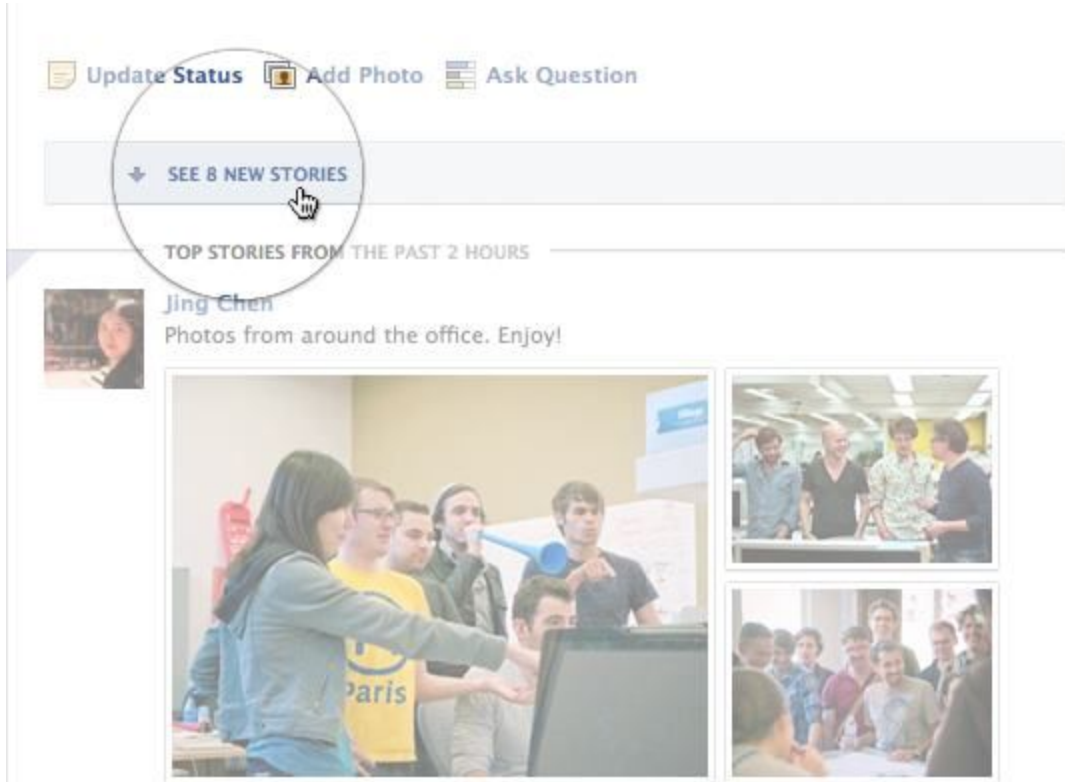


Image 11: Screenshot from Facebook's News Feed update in 2011 (via Wayback Machine).

This was a significant shift from reverse chronological feeds, which other major social media platforms, like Twitter, were using at the time. Algorithmic curation became based on metrics like virality (how many times a post is “liked” or shared) and our “preferences” (based on data points that capture our interactions and behavior). In a way, algorithmic curation solved a problem (or perhaps, the intentional choice) that Facebook created with the original introduction of the News Feed: the hypersaturation of information on social media. As social networks grew and the News Feed began to function as a place for also getting traditional “news”, it became increasingly impossible to see all of the content that passed through our feeds.

Twitter maintained a reverse chronological feed as a default several years longer than Facebook -- until 2016. Perhaps this is because browsing content posted in real-time has been a key part of Twitter's brand, reflected in their early slogan, "See what's happening — right now"⁹² (which has only been slightly modified in 2019 to "See what's happening in the world right now"). Another key aspect of Twitter's brand was about giving users control, through features like "quiet times"⁹³, a setting for turning off Twitter updates for certain periods of time. In many ways, Twitter has explicitly differentiated itself from Facebook by maintaining the chronological feed and emphasizing its features for user control.

Therefore, when Twitter decided to introduce an algorithmically curated feed in 2016, Twitter users collectively expressed their lament over the change, even leading to the viral hashtag, #RIPTwitter⁹⁴. Many users expressed anger, frustration, and surprise over this shift from a chronological feed to an algorithmically curated one.



⁹² Source: <https://web.archive.org/web/20081030092504/http://search.twitter.com/>

⁹³ Source: <https://web.archive.org/web/20081031021416/http://twitter.com/help/how>

⁹⁴ Source: <https://web.archive.org/web/20160207110712/https://twitter.com/hashtag/RIPTwitter?src=hash>

*Image 12: Twitter user, @CulturedRuffian, shared a GIF expressing anger over Twitter’s new feed algorithm.*⁹⁵

Since Twitter introduced this algorithmic change, users began sharing ways to “avoid the algorithm” -- specifically to avoid Twitter’s algorithmically curated feed. Not long after Twitter introduced the change, users like @xeenarh shared, “Creating a list with all I’m following to avoid the algorithm & enjoy what USED TO BE the @twitter experience feels like an act of rebellion”⁹⁶ (Image 14). The “list” that @xeenarh is referring to Twitter’s “List” feature⁹⁷, which allows users to curate a group of Twitter accounts. Users can then browse a List feed, or “timeline”, which only shows content from accounts a user added to the List.



Image 13

@xeenarh’s “act of rebellion” reflects a broader range of anti-algorithmic tactics that several social media users have taken to “avoid” and subvert algorithms for content curation. Anti-algorithmic tactics have emerged not only in the aftermath of algorithmic changes made by Twitter but also against the backdrop of increasing criticisms of social media platforms more broadly. Major platforms like Facebook and Twitter, once celebrated for fostering social connections and for enabling mass civic movements⁹⁸, have since been blamed for enabling a

⁹⁵ Source: <https://twitter.com/CulturedRuffian/status/696131512278732800>

⁹⁶ Source: <https://twitter.com/xeenarh/status/710723495710347265>

⁹⁷ See <https://help.twitter.com/en/using-twitter/twitter-lists>

⁹⁸ Hwang and Kim, “Social Movements.”

range of socially undesirable effects, from ideological echo chambers⁹⁹ to the viral spread of false news.¹⁰⁰

While many social media users may recognize negative aspects of social media at both individual and collective scales, I am especially interested in how some users have actively counteracted these aspects -- specifically, aspects related to algorithmic content curation. In order to explore this, I interviewed users of Gobo, a social media tool aimed at allowing users to understand and “control the algorithms” on social media feeds.¹⁰¹ I was particularly interested in how Gobo users were engaging with anti-algorithmic tactics, given their interests in algorithms and active experimentation with alternative social media tools.

Gobo was started in 2017 by researchers at the MIT Center for Civic Media in response to issues raised around algorithmic filtering on social media.¹⁰² It “aims to help users control what’s hidden from their feeds, add perspectives from outside their network to help them break filter bubbles, and explore why they see certain content on their feed.”¹⁰³

⁹⁹ Pariser, *Filter Bubble*.

¹⁰⁰ Lazer et al., “Fake News.”

¹⁰¹ See <https://gobo.social/about>.

¹⁰² Zuckerman, “Gobo.social.”

¹⁰³ Bhargava et al., “Gobo.”

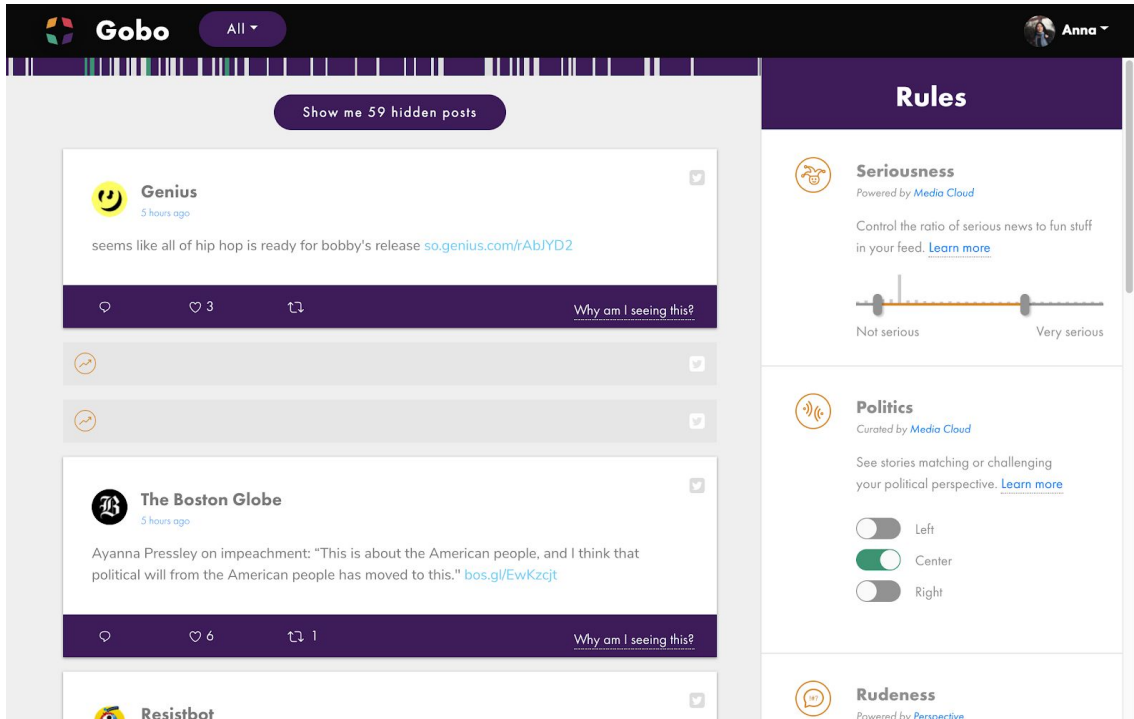


Image 14: Screenshot of Gobo on desktop.

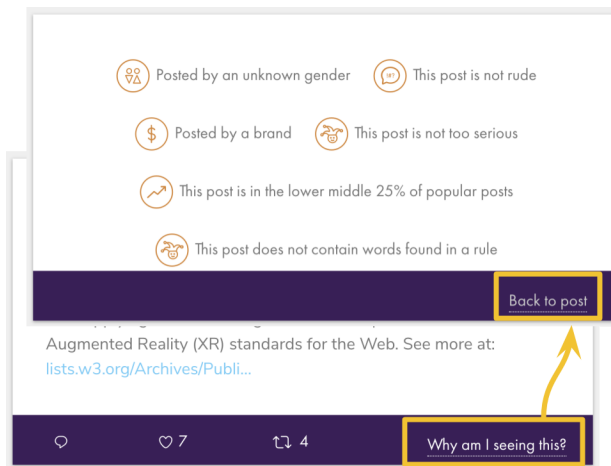


Image 15: Each post on a Gobo user's feed features a "Why am I seeing this?" button. Clicking this reveals the "back of the post," which contains brief explanations of how the post was classified by each rule.

I joined Gobo as a designer in 2018 as we planned to redesign the interface and understand how Gobo users were engaging with the tool. As we aimed to better understand user engagement, I was simultaneously interested in Gobo users' experiences with algorithms on social media more broadly -- and specifically, what kinds of anti-algorithmic tactics they had adopted. In this case study, I focus on the following questions:

1. What kinds of frustrations have Gobo users experienced on algorithm-driven social media feeds?
2. What tactics have Gobo users adopted to avoid algorithmic content curation? How are these tactics informed by frustrations and folk theories around algorithms?

Methods

In collaboration with Gobo's lead developer, I conducted semi-structured interviews with 10 Gobo users, who indicated they were willing to participate in conversations in an earlier research survey sent out to all Gobo users.¹⁰⁴ Interviewees were asked to fill out an online pre-survey to collect demographic information (Image 16). 9 identified as men and 1 identified as a woman. Ages ranged from 18 to 75+ years, and locations spanned across four continents: Africa, Asia, Europe, and North America. These interviewees represent an especially proactive subset of users when it comes to experimentation with new tools, even compared to Gobo users in general -- who already might be seen as more willing to adopt new tools than the average social media user. Most interviewees were especially active in exploring open-source online tools and engaging in conversations around social media. They presented strong opinions and experiences around social media algorithms and, oftentimes, took steps to address their frustrations around these algorithms.

¹⁰⁴ Bhargava et. al, "Gobo."

Interviews were conducted via video or phone and lasted between 40 minutes to 1 hour. The first portion of the interview focused on the interviewees' background and experiences with social media platforms, and the second portion of the interview focused more on particular frustrations or satisfaction with these platforms.¹⁰⁵ Following the interviews, I conducted a qualitative analysis of the interview transcripts to code the data and develop themes. For this case study analysis, I present themes around frustrations with algorithm-driven feeds and examples of anti-algorithmic tactics interviewees had shared.

Occupation [label]	Age group (years)
Lab director [P01]	45-54
Freelance data scientist [P02]	35-44
Retired auditor [P03]	75+
Communications consultant [P04]	55-64
Content strategist [P05]	45-54
CG artist [P06]	35-44
Research assistant [P07]	18-24
Software engineer [P08]	25-34
Student [P09]	18-24
Librarian [P10]	65-74

Image 16: Demographic information of 10 interviewees.

¹⁰⁵ See Appendix for full interview script.

User frustrations with algorithm-driven social feeds

During the interviews, I probed into the interviewees' experiences with browsing algorithm-driven social media feeds. In addition to asking interviewees what motivated them to try out Gobo, I also asked them about their experiences with other social media platforms. Interviewees shared their opinions on several popular platforms, including Twitter, Facebook, Tik Tok, and Pinterest. After analyzing interview transcripts, I specifically identified three common frustrations around algorithm-driven feeds: a lack of diverse perspectives, passive consumption, and infinite streams of content.

Lack of diverse perspectives

"Facebook kind of locked me into an echo chamber." - P04

Five interviewees specifically talked about their experiences with browsing Facebook when discussing their frustrations with algorithmically curated content. In the U.S., particularly following the 2016 presidential elections, critiques of Facebook as an echo chamber of political content¹⁰⁶ became increasingly prominent. P04 alluded to this time period when he described the shift he noticed in the diversity of political opinions on his Facebook feed: *"I see very little stuff that challenges my preconceived ideas and challenges my biases. I would actually like to see a little bit more. And I noticed a big drop off in that kind of stuff years ago. Like about two three years ago."* P04 expressed a desire to see more content that "challenges" his views and found that Facebook's News Feed has actively evolved in doing the opposite.

In response to the lack of diverse perspectives seen on Facebook, another interviewee shared his friend's experience with trying to "adjust" Facebook's algorithm: *"I guess that algorithm of Facebook is too strong that he had to try very, very hard in order to adjust this*

¹⁰⁶ Hern, "Filter bubbles."

algorithm to be incorporating different people, different opinions, from the other side” (P08). In this case, P08 explains a folk theory around Facebook’s News Feed algorithm as intentionally pushing more homogenous content into the feed -- the lack of diverse perspectives on their Facebook feed is directly related to the algorithm. However, this folk theory also suggest sthat Facebook’s News Feed algorithm is possible to “adjust” over time to incorporate perspectives “from the other side.”

While this user aimed to expand the political diversity of content by working within the platform the “adjust this algorithm”, other interviewees, such as P10, engaged with other platforms, such as Google News. P10 specifically appreciated Google News’ algorithmic content curation as being “sensitive to what I want to see” -- in contrast to Facebook’s algorithmic content curation which is “too strong” and difficult to adjust. However, P10 still wanted to engage with diverse political perspectives on social media as well and expressed interest in other platforms that can help him “avoid living in a bubble.”

Being able to see a variety of perspectives is an important feature of social media for these interviewees, and on platforms like Facebook, the algorithm is seen as actively inhibiting these perspectives. For one interviewee, attempting to work within Facebook and its algorithmic logic has been one way to incorporate “different opinions”, while another interviewee’s approach was to engage with other platforms and their curation algorithms. In both cases, the lack of diverse perspectives on algorithm-driven social media feeds has been a source of frustration that has motivated alternative forms of engagement.

Passive consumption vs. active participation

“I think I enjoy like the process of just having a little bit more control and looking for things myself, like looking for interesting things myself.” - P09

Four interviewees expressed wanting “more control” over their browsing experience on algorithm-driven feeds. P09 specifically described how they feel when browsing platforms like Facebook: *“I think I'm just kind of like turned off. Like, my brain isn't really engaging a whole lot when I'm doing that, and I'm just kind of like passively browsing through content rather than actively looking for it. Which I think is less interesting and less engaging”* (P09). Social media platforms have used algorithms, in part, to reduce effort for the user, but for users like P09, “passively browsing” has also had negative effects of being “less interesting and less engaging.” While algorithms might aim to surface more interesting and engaging content, the browsing experience may become “less interesting and less engaging” for a user who no longer has to actively participate in the content discovery process.

However, algorithms can be seen as having a positive effect on the browsing experience when they are incorporated with user-driven content discovery. P09 specifically mentioned Pinterest as providing a happy medium between algorithm-driven and user-driven discovery: *“I think it has a very nice balance between being guided by an algorithm and recommendations and having your own like self-driven process.”* For P09, algorithms that mostly eliminate the need for active user participation can make the browsing experience passive and unengaging. He emphasizes the importance of having ownership and input in the browsing process, which should be supported by but not fully driven by algorithms.

Addictive, endless feeds

“You can spend a lot of time on it, like gambling, or it's like drinking alcohol. Basically you're not getting anything out of that. You're not learning new things.” - P07

A commonly shared frustration amongst interviewees is the infinite stream of content that is surfaced in algorithm-driven social media feeds, such as TikTok. P07 described TikTok and

its video feed: *“That’s a very toxic, addictive app. Basically short videos, and there’s no search. You just see one video after another”* (P07). In recent years, TikTok has gained immense popularity, and the company that developed the app, ByteDance, has notably been described as an AI company rather than a social media company¹⁰⁷. P07 points out that TikTok lacks a “search” functionality, further narrowing what users can see on the app to content that is pushed by TikTok’s algorithm, which curates content based on past videos a user engages with.¹⁰⁸

Frustrations around infinite content streams were echoed by P01, who lamented that “there’s always more” content than he can possibly consume. Reflecting on his experiences with browsing content on social media platforms, P01 shared, *“I haven’t really gotten in the habit of ‘finish this, on to the next, finish this and on to the next’.”* Because new recommendations are always being put in front of him, he feels like content consumption is always ongoing. *“If any of these gave me a sense of closure, then maybe I would cycle through them better”* (P01). On the most widely used social media platforms, endless, algorithm-driven feeds have become a key feature for engaging users -- but they have also been a source of frustration for users who feel a lack of closure and control.

Anti-algorithmic tactics for browsing social media content

As interviewees discussed their experiences with browsing content on various social media platforms, they shared several frustrations around algorithm-driven social feeds. In part, they discussed these frustrations as part of their motivation for signing up for Gobo. For example, P10 hoped that Gobo would help them incorporate more diverse perspectives into their Facebook feed. While some interviewees specifically talked about frustrations with particular algorithm-driven feeds like Facebook at TikTok, other interviewees spoke more

¹⁰⁷ Pham, “ByteDance.”

¹⁰⁸ Ibid.

broadly about the negative effects of endless feeds and passive browsing on their social media experiences. In some cases, frustrations were informed by folk theories of algorithms -- as was the case with P08's friend, who believed that Facebook's News Feed algorithm intentionally excluded diverse perspectives on their Facebook feed -- and motivated attempts to *adjust* the algorithm. In other cases, frustrations further motivated anti-algorithmic tactics by users to *avoid* algorithms on their social media feeds.

Interviewees specifically shared tactics for avoiding content curation algorithms, both through using features within a platform and experimenting with alternative tools and platforms for social media browsing. These anti-algorithmic tactics primarily focused on content curation on Twitter and Facebook, and I examined how tactics are used and how they are motivated or related to frustrations around algorithm-driven feeds.

Self-driven curation (via platform feature)

Multiple interviewees engaged in tactics to avoid content curation algorithms entirely by taking on the curation process themselves. In particular, four interviewees shared about their use of Twitter's "List" feature to curate groups of accounts they were most interested in reading or groups of accounts separated by subject. P05 shared, *"I implemented a personal tactic to avoid [the feed algorithm] by creating a Twitter List of just the accounts that I really want to read. And that gives me everything in chronological, so I actually avoid the front Twitter homepage and their algorithm on that page."* P05 explicitly describes his use of Twitter Lists as a way to avoid Twitter's feed algorithm -- to drive his own content curation and focusing on seeing things he "really wants to read."

Other interviews used Twitter Lists both to avoid Twitter's algorithm and to have the option to browse content separately by different domains. P06 shared, *"I've been trying to use*

Twitter lists to separate the 3D stuff from the political stuff.” As a CG artist, P06 expressed that they enjoyed using Twitter to see what people were sharing in the 3D modeling field as well as to keep up with global political news. However, they wanted a better way to use Twitter that would support their different motivations for browsing and began to experiment with Twitter Lists as a browsing tactic. P02 expressed a similar desire and shared, *“It would be nice to have, like all of my music related accounts that I follow and all the AI and NLP stuff in one area, so that I can sort of decide if I want to consume that at some point or another.”*

The use of Twitter Lists as an anti-algorithmic tactic may reflect a desire for greater control over content curation on the platform. Though Twitter Lists had been viewed as a more niche feature “largely adopted by Twitter power users”, the feature was redesigned in 2019 to be made more prominent and easier to access in Twitter’s interface -- perhaps acknowledgement of the feature’s importance from the company itself. Given the frustrations described around passive consumption and lack of diverse perspectives on social media platforms, Twitter Lists may offer a way for users to drive their own content curation while being able to access the variety of content and communities on Twitter.

Self-driven filtering (via third-party plugin)

While Twitter Lists may offer a way to “avoid the algorithm” and enable greater control on the Twitter browsing experience, other platforms, such as Facebook, lack features that enable this tactic within the platform. As a result, some interviewees have experimented with tools outside of the platform that allow them to hide undesired content surfaced by the platform’s algorithm. P04 specifically described how they use a plugin called *Social Fixer*, a tool started by independent developer, Matt Kruse. On its website, *Social Fixer* lists key features of the plugin,

including the ability to “filter your news feed by keyword, author, and more” and to “auto-switch to the most recent news feed.”¹⁰⁹

P04 shared a positive experience with using *Social Fixer* to modify their algorithm-driven Facebook feed, explaining that it “blocks ads and blocks stuff that annoys me.” For P04, *Social Fixer* enables them to customize their browsing experience rather than having it customized by Facebook and its algorithms. Through offering users a way to take a more active role in the browsing experience, a tool like *Social Fixer* may address frustrations raised around passive consumption on algorithm-driven feeds. Furthermore, by using *Social Fixer*, P04 theorizes that they are not only gaining control in customizing their feed but also in “subverting the business model of Facebook.” P04’s use of *Social Fixer* may be seen as an anti-algorithmic tactic which allows him to modify the outputs of Facebook’s News Feed algorithm and filter algorithm-driven content on his feed. Because *Social Fixer* is a third-party plugin, it may enable more flexibility and customizability than an in-platform feature like Twitter Lists. However, it is also more vulnerable to changes in Facebook’s code and policies, as other online users of *Social Fixer* have noted issues with the plugin¹¹⁰ following changes to Facebook’s feed.

Limited curation (via third-party app)

A different anti-algorithmic tactic noted by four interviewees involved a third-party app called Nuzzel¹¹¹. Nuzzel is described as an app that curates “a feed of news stories shared by friends on Twitter, Facebook, and other social media” and sends this curated set of stories to a user’s inbox.¹¹² Interviewees specifically mentioned using Nuzzel as a way to keep up with Twitter content without having to browse their Twitter feeds, which present an endless,

¹⁰⁹ See <https://socialfixer.com/>.

¹¹⁰ Kruse, “Social Fixer.”

¹¹¹ See <https://nuzzel.com>.

¹¹² Edwards, “Amazing New App.”

overwhelming stream of content -- a previous frustration expressed by several interviewees. Rather than consuming content via Twitter's algorithm-driven, limitless feed, interviewees who used Nuzzel preferred seeing a limited set of content shared by their friends.

P10 expressed that they prefer the content curated on Nuzzel, rather than content on their algorithm-driven Twitter feed, because "it's mediated by other people whose judgment I trust." To an extent, however, P10's understanding of Nuzzel may present a folk theory around how Nuzzel works -- as being driven by human curation rather than algorithmic curation. However, although Nuzzel may not algorithmically personalize content in the same way that Twitter does, Nuzzel likely employs algorithms to present a limited set of content from a user's network.

Nuzzel has been described on Business Insider as a "way to fixing Twitter's chaotic news-feed problem."¹¹³ The app presents an alternative form of content curation to the endless feed paradigm used by platforms like Twitter and Facebook -- curation that is limited and based not on a user's personal browsing behaviors but, rather, based on what their network is sharing. While Nuzzel itself may employ algorithms to determine the limited set of stories to show a user, I argue that the interviewees' use of Nuzzel can be seen as an anti-algorithmic in that it is aimed at avoiding *Twitter's* algorithms. Nuzzel offers a different paradigm for content curation and browsing that subverts the dominant paradigm used by Twitter -- algorithm-driven feeds that optimize for attention and engagement.

Conclusion

¹¹³ Ibid.

The interviews with Gobo users offer a lens into frustrations around algorithm-driven social media feeds and the tactics that users have adopted to “avoid” or subvert content curation algorithms. Anti-algorithmic tactics were motivated by frustrations with social media feeds but also, in part, by folk theories around how content curation algorithms work. In addition, the anti-algorithmic tactics adopted by Gobo users may reflect the culture of experimentation and early adoption of tools amongst the broader open source community. In contrast with the previous case study, in which anti-algorithmic tactics extended practices in the online Black community, the anti-algorithmic tactics of Gobo users extend practices of the open source community, which several interviewees actively engage with.

In addition, these anti-algorithmic tactics may reflect the ways in which users have sought to gain agency in their browsing experiences and the ways that platforms limit this agency. As such, anti-algorithmic tactics have been taken up both within platforms and outside of platforms via third-party tools. These tactics address various frustrations around the lack of diverse perspectives, passive consumption, and endless streams of content on social media platforms. Through adopting various anti-algorithmic tactics on platforms like Twitter and Facebook, users have been able to assert their own preferences in content browsing while continuing to stay connected with news and updates from their online communities.

While these anti-algorithmic tactics may be aimed at specific aspects of platform algorithms, I argue that, in general, interviewees were not opposed to the use of algorithms in content browsing in general. When expressing his frustrations over the lack of active participation in content discovery, P09 suggested that discovery can be enhanced by “an algorithm and recommendations” as long as it is balanced with a self-driven process. The use of tools like Nuzzel by multiple interviewees may also indicate interest in content curation

algorithms that simply operate under a different set of parameters, based on surfacing popular content in a user's network rather than optimizing based on a user's past behavior.

User engagement with anti-algorithmic tactics suggest opportunities for co-algorithmic content curation, in which algorithm-driven curation is not eliminated altogether but is used in conjunction with user-driven curation. Co-algorithmic curation could enable users to better tailor algorithms to what they are actually interested in and to use algorithms to hide irrelevant content -- similar to how P04 used *Social Fixer* on their Facebook feed. However, I argue that co-algorithmic curation requires further user knowledge of and participation in algorithm-driven curation. As such, how might algorithmic systems be designed to empower users with more explicit understanding and control?

Chapter 3

Designing for user agency on algorithmic platforms

Examining anti-algorithmic tactics reveals various frustrations and negative experiences users have had on algorithm-driven social media platforms. In the first case study, I examined how Facebook users developed folk theories around the platform's content moderation algorithms, based on experiences with flagged content and account bans. These folk theories informed the anti-algorithmic tactic of using alternative spellings to avoid detection by Facebook's content moderation algorithms. In the second case study, I investigated the anti-algorithmic tactics of Gobo users, informed by frustrations and folk theories around browsing content on algorithm-driven social media feeds. While these case studies presented distinct anti-algorithmic tactics, informed by different motivations and used for different purposes, they present some common ways that algorithmic platforms limit user knowledge and agency.

In the first case study, varying folk theories emerged in part because of Facebook's lack of information around content moderation on their platform. Users whose posts were removed or were given account bans were not given clear reasons by Facebook itself and, thus, had to develop their own reasonings for how content moderation decisions were being made. In addition, Facebook does not make known which content moderation decisions are made by algorithms, where these algorithms are used on the platform (i.e. on public posts vs. private groups), and what guidelines these algorithms enforce. Furthermore, though Facebook has announced plans to establish an Oversight Board¹¹⁴, which would allow users to appeal content

¹¹⁴ Harris, "Facebook's Oversight Board."

moderation decisions, there is currently no clear way for users to specifically appeal algorithm-driven decisions or otherwise inform algorithmic decision-making.

In the second case study, Gobo users employed anti-algorithmic tactics to avoid content curation algorithms and to take on a more active role in content browsing. Several Gobo users talked about how their algorithm-driven feeds lacked diverse perspectives and how they found content browsing more interesting when they could actively participate in the process rather than passively consume content. On Facebook and Twitter, users found that there were few ways to change algorithm-driven feeds within the platform and resorted to using third-party tools to gain more agency over their content browsing experience. These platforms, again, provide little information on how algorithms determine that content that is shown on a user's feed and, moreover, few ways to customize an algorithm-driven feed.

Both case studies illuminate issues around algorithm-driven social media that can emerge from a lack of transparency -- limited knowledge of how an algorithm works -- and a lack of agency -- limited opportunities to inform and participate in content curation and moderation. In recent years, alternative social media platforms have emerged which respond to issues raised around algorithm-driven social media, such as Are.na and Mastodon. Are.na is a platform where users can collect content from across the web and "connect ideas with other people by collaborating privately or building public collections for everyone."¹¹⁵ Notably, Are.na touts that their platform contains "no ads, likes, or recommendations"¹¹⁶ and has been described as an "anti-algorithmic" platform.¹¹⁷¹¹⁸ On Are.na, users drive content curation through collecting their own content, making their own connections, and discovering content through active searches and shared content in their network.

¹¹⁵ Are.na, "About."

¹¹⁶ Ibid.

¹¹⁷ Bezic, "Anti-Algorithmic Networks."

¹¹⁸ As an amusing side note, Are.na also sells a "F*** an algorithm" mug in their online store.

On the other hand, as an alternative to algorithm-driven content moderation, Mastodon presents a way for smaller social networks to determine and enforce content moderation guidelines. Similarly to Are.na, Mastodon advertises itself as an algorithm-free, ad-free platform: “Your feed is chronological, ad-free and non-algorithmic—you decide who you want to see!” Mastodon is further described as a “decentralized” social media network¹¹⁹; rather than joining a centralized Mastodon network, users join “instances” -- subnetworks that any user can host on private servers. Instances each have different sets of content moderation guidelines, determined by moderators and, in some cases, users of an instance. On Mastodon instances, content moderation is determined and enforced not by algorithms but by users and moderators.

While Are.na and Mastodon may present ways to address issues with algorithm-driven social media, as platforms, they take an explicitly anti-algorithmic stance that positions themselves against social media platforms like Facebook and Twitter. However, considering that the majority of social media users use these algorithm-driven platforms -- and that algorithmic systems are increasingly present in other digital experiences -- I argue that it is also important to consider how these algorithmic systems might not be eliminated, but rather, designed differently. In examining anti-algorithmic tactics in the previous case studies, I found that users are not necessarily *anti-algorithm* but may engage with anti-algorithmic tactics because they lack knowledge and agency on algorithm-driven platforms. Driven by insights from these case studies, I examine two design frameworks -- seamful design and contestability -- to consider how algorithmic systems can better support user awareness of algorithms and participation in algorithmic decision-making.

¹¹⁹ Glaser and Oremus, “New Social Network.”

Seamful design: enhancing user understanding of algorithms

Anti-algorithmic tactics are motivated, in part, by negative perceptions and frustrations around algorithms. In some cases, this may reflect limited understandings around how algorithms work due to a lack of algorithmic transparency. In fact, DeVito et al. argue that “limited algorithmic transparency can cause users’ dissatisfaction, lower users’ trust in the system, and sometimes lead the user to stop using the platform.”¹²⁰ In the first case study, a lack of transparency around Facebook’s content moderation algorithms led to varying folk theories of how they work and a general mistrust in how content is moderated. Naher et al. further assert, “As end-users are the central actors in online social systems, from the trust and transparency perspective of the user, it is important to design for reducing the opacity of the algorithm used in the moderation decision making.”¹²¹

Algorithmic systems may intentionally obscure how algorithms work or otherwise assume user knowledge, which can drive negative perceptions around algorithms and motivate anti-algorithmic tactics. Karahalios et al. point out that recent work in HCI has “abandoned the idea of an accurate mental model”, which assumes that a user has a clear picture of how a system operates; instead, “more recent work has shifted toward the signifiers that the system provides that allow a user to construct their understanding of it, which is always partial.”¹²² In algorithmic systems, providing clear signifiers may be particularly necessary in enabling users to gain fuller understandings of how the algorithms work and what effects they have on the system.

Seamful design offers one framework for thinking about how these signifiers can be incorporated in algorithmic systems. Unlike *seamless* designs of algorithmic systems, in which

¹²⁰ Naher et al., “Algorithmic Understandability.”

¹²¹ Naher et al., “Algorithmic Understandability.”

¹²² Hamilton et al., “Algorithm Awareness”, 638.

“the black box is sealed or even invisible”¹²³, *seamful* designs “emphasize mechanism” and “makes system infrastructure elements visible when the user actively chooses to understand or modify that system.”¹²⁴ On social media feeds, I argue that seamful design can be particularly important in enabling algorithmic awareness, given the lack of knowledge around how algorithms influence a user’s feed. While incorporating seams into a social media feed may not directly tell a user how an algorithm works, they “invite the user to explore and discover connections in the system through manipulation, comparison, and feedback.”¹²⁵

In their research on user browsing experiences with Facebook, Eslami et. al created a system called FeedVis to incorporate “seams” in Facebook’s News Feed design to bring “visibility” to Facebook’s hidden feed algorithm. FeedVis “highlights the content that the algorithm excluded from display and reveals social patterns by disclosing whose stories appeared and whose were hidden in News Feed.”¹²⁶ FeedVis was used by Eslami et al. in their research study to evaluate user understandings and perceptions around Facebook’s News Feed algorithm.

Meanwhile, in my work as a designer for Gobo, we drew inspiration from FeedVis and applied the concept of seamful design to enable more algorithmic transparency on Gobo’s feed. On Gobo, users can apply different algorithms (called “Rules”) to their social media feeds to hide certain kinds of posts. In Gobo’s original interface, posts were completely removed from the feed when algorithms were applied. As a result, it wasn’t immediately apparent to users when posts were being hidden and where gaps in their feed existed. Applying the framework of seamful design, we decided to redesign the interface by visually collapsing posts when they were hidden by user-controlled algorithms -- instead of removing them from the feed entirely --

¹²³ Hamilton et al., “Algorithm Awareness”, 633.

¹²⁴ Eslami et al., “Folk Theories”, 2373.

¹²⁵ Ibid, 2373.

¹²⁶ Eslami, “Hidden Algorithms”, 58.

and providing explanations of how a post was evaluated by each algorithm. In addition, we added a visualization bar to the top of the feed which indicated where posts were both hidden and added to the feed. By exposing the gaps or “seams” in the feed, we hoped to make more transparent how Gobo’s algorithms were being applied to a user’s feed (Image 17).

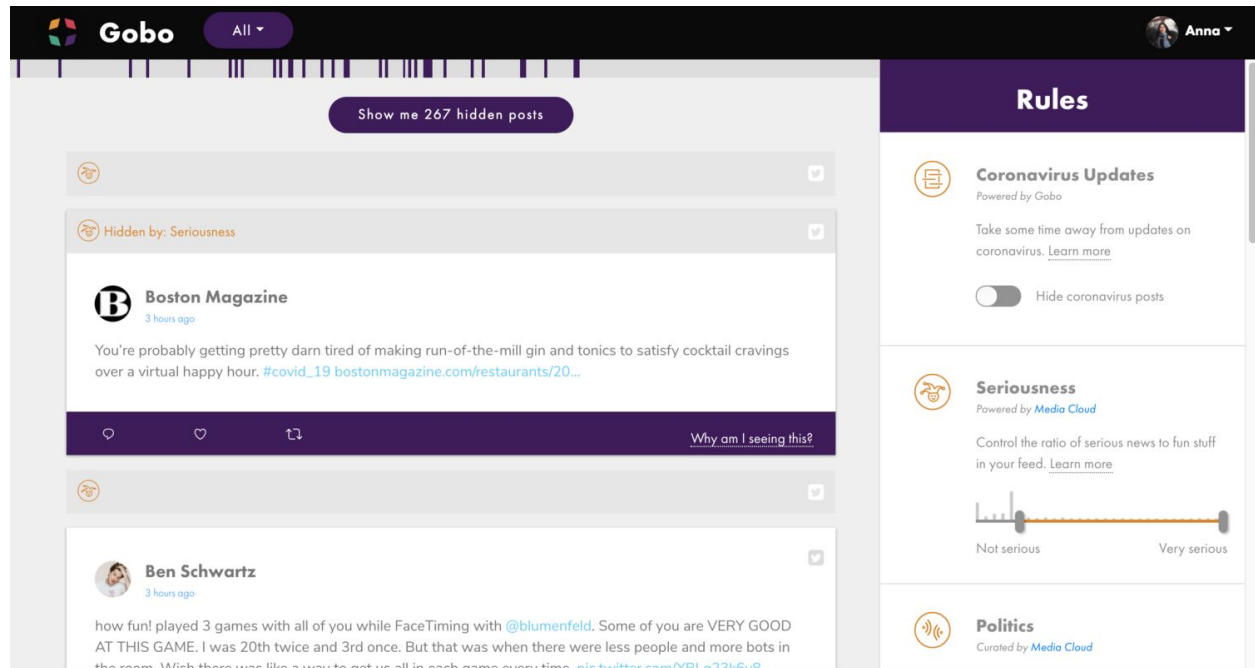


Image 17: Screenshot of Gobo. Posts hidden by different “Rules” are collapsed on the feed, and a visualization bar at the top shows where posts were hidden throughout the feed.

Seamful design provides a framework for thinking through how algorithmic systems can be designed to promote more transparency and awareness around algorithms. Seamful design can help to build more accurate user understandings of an algorithmic system, which “might lead to increased user agency and success in achieving goals, as well as increased user trust in these systems.”¹²⁷ A seamful interface can also “be seen as a tool for the user to be empowered

¹²⁷ DeVito et al., “Algorithms Ruin”, 3163.

against a system when their interests might differ”¹²⁸ and enable understanding around “potential biases” that algorithms might introduce.¹²⁹

Contestability: enabling user participation in algorithmic decision-making

While seamful design may raise awareness and understanding around how algorithms work, “contestability” can motivate the design of algorithmic systems in which users can more directly participate -- and when necessary, intervene -- in algorithmic decision making. In addition to engaging with seamful design in their research, HCI researcher Motahhare Eslami further highlighted contestability as a design principle for “allowing users to voice their arguments and disagreement, and appealing algorithmic decisions – via a form of human-algorithm dialog.”¹³⁰ Contestability may be especially important where algorithmic decision making can have unintended negative impacts on users -- such as account bans experienced by Black Facebook users. In these cases, it becomes important not only to know how algorithmic decisions are being made but also to challenge, or “contest”, these decisions. As Eslami states, “knowing the reasons behind an algorithmic decision without the agency to affect it makes users feel powerless.”¹³¹

While contestability is easier to incorporate in algorithmic systems with lower stakes and more straightforward functionality, complex systems with various actors, such as social media, may require ways for users to present further justification for their disagreements with algorithmic decisions. In their work on contestability that emerged from working on an automated training and assessment tool for psychotherapists, Hirsch et al. argue that appealing more complex algorithmic decisions could require users to “marshal evidence and create

¹²⁸ Eslami et al., “Folk Theories”, 2380

¹²⁹ Eslami, “Hidden Algorithms”, 60.

¹³⁰ Eslamimehdiabadi, “Participating and Designing around Algorithmic Socio-Technical Systems.”, 137.

¹³¹ Ibid, 137.

counter narratives that argue precisely why they disagree with a conclusion drawn by an AI system.”¹³² For example, appealing algorithm-driven content moderation decisions on Facebook may require users to provide cultural and social context around content that has been shared.

Hirsch et al. outline several strategies for designing algorithmic systems for contestability. Among these strategies is to provide “mechanisms for users to ask questions and record disagreements with system behavior” and to look for “aggregate effects that may not be apparent to individual users.”¹³³ While it is important to enable users to interrogate an algorithmic system themselves, certain biases in algorithmic systems may only be “made visible by looking at behavior across multiple users and sessions”¹³⁴ and thus require designers of a system to be more intentional in identifying -- and correcting for -- discriminatory behaviors. While contestability may be relevant for enabling more user-driven algorithmic content curation, it is especially important for thinking through algorithmic content moderation, which can have significant impacts on what people can both say and see online.

In thinking about how contestability might be incorporated into an algorithmic system, I return again to my work with Gobo and how transparency around algorithmic evaluations may enable contestability. In particular, Gobo provides an alternative feed view that only shows posts that are “hidden” from the user’s home feed (Image 18). In the alternative feed view, posts are explicitly labeled with the algorithms they are hidden by. This may enable a user to interrogate how algorithms are evaluating content on their feed and more easily identify inaccurate evaluations. On Gobo, the alternative feed view can serve as a “mechanism for users to ask

¹³² Hirsch et al., “Designing Contestability”, 98.

¹³³ Ibid, 98.

¹³⁴ Ibid, 98.

questions” around algorithmic evaluations and, subsequently, motivate users to appeal or contest these evaluations.

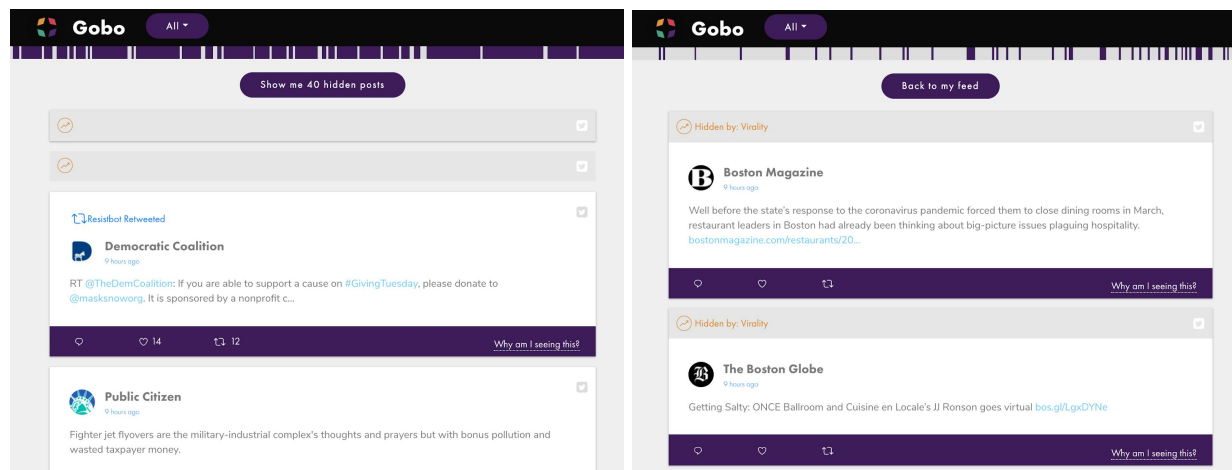


Image 18: Screenshots of Gobo. Gobo has a home feed and an alternative “hidden posts” feed. In the alternative feed, only posts that are hidden from the home feed are shown and are labeled with the particular algorithms they were hidden by.

Work around contestability in algorithmic systems is relatively nascent but offers an interesting path forward in thinking through user-centered designs of algorithmic systems. As Eslami states, contestability “can improve users’ perceptions of fairness and accountability of algorithmic systems.”¹³⁵ In enabling users to not only better understand algorithms but also to provide input into algorithmic decision-making -- particularly when biased or incorrect decisions are produced -- contestability could enable greater user agency on algorithmic systems.

¹³⁵ Eslamimehdiabadi, “Algorithmic Socio-technical Systems”, 137.

Conclusion

Algorithms are increasingly embraced -- and even touted -- by online platforms which aim to optimize for user engagement. Algorithms have become unique assets of companies like ByteDance, which has created an “algorithmic infrastructure” to power TikTok and dozens of other online content sharing platforms.¹³⁶ As companies incorporate more and more algorithms into their products, users’ understandings of and relationships with algorithmic systems are evolving. In recent years, algorithms have entered more widespread online discourse as users grapple with algorithmic influence on popular online platforms such as Facebook and Twitter. While algorithms may quietly exist in the background of many users’ online experiences, some users are increasingly frustrated and concerned with how algorithms are shaping what they can see and say online.

Anti-algorithmic tactics have become a way for users to actively subvert -- or otherwise call attention to -- algorithmic systems which may actively marginalize or otherwise have negative effects on users. Personal experiences with these algorithmic systems have led users to express frustrations around algorithms and folk theories of how they work. In part, anti-algorithmic tactics are motivated by these frustration and folk theories. However, they are also shaped by existing practices that users engage with in their online communities, such as the use of alternative spellings in Black online communities and the adoption of new tools in open source communities.

In this thesis, I explored two case studies of social media users who have engaged with anti-algorithmic tactics. By probing into these tactics, I discovered different understandings around and responses to algorithm-driven content moderation and content curation. In the first

¹³⁶ Le, “Bytedance.”

case study, examining Facebook users' explanations behind alternative spellings revealed negative perceptions around Facebook's content moderation algorithms and folk theories around how these algorithms work. In the second case study, probing into the frustrations and tactics of Gobo users revealed the lack of agency that these users feel when browsing content on platforms like Twitter and Facebook. These case studies offer a window into some of the pain points users have with algorithmic systems and the broader ways in which these systems limit user knowledge and participation. Anti-algorithmic tactics may enable some users to address issues with algorithmic systems and, at scale, could eventually reshape how algorithmic systems work. However, algorithmic systems can also be critically reassessed and designed in ways that might better support user understanding and agency.

Seamful design and contestability are two approaches for thinking through how algorithmic systems could be designed to enable user participation. These design approaches can be considered within a broader conversation around the *ethical design* of algorithmic systems. As social science researcher Faye Miller writes, "Ethical design means users must be able to define their relationship with a platform — not the other way around."¹³⁷ But how might we get to ethically designed social media? As a Gobo interviewee expressed, "It would be nice if the Internet wasn't this massive thing and wasn't completely co-opted by large companies but I don't see that changing very soon. So yeah, an alternative would be cool, but that seems difficult to imagine."

Given the financial incentives that social media companies have in maintaining opaque algorithmic systems, wide scale change in these systems may be difficult to imagine. However, as researchers, journalists, and social media users have raised issues around algorithms to the public, companies have been pressured to begin adopting changes in their algorithm-driven

¹³⁷ Johnson, "Ethical Social Media."

platforms. For example, in 2019, Facebook introduced a feature -- called "Why am I seeing this post?"¹³⁸ -- which supports greater algorithmic transparency on their News Feed. While these changes might seem small, I believe that they reflect the potential for more transformational changes in the ways algorithmic systems are designed.

This thesis explored anti-algorithmic tactics in order to understand how social media users are grappling with algorithmic systems. As algorithms become increasingly influential not only in what we see online but also in shaping our lives offline, I argue that we need to continue examining these algorithmic systems and our relationships to them. As users call algorithms into question, we might begin to challenge algorithmic power and find ways to reclaim and redesign algorithmic systems.

¹³⁸ Sethuraman, "Why Am I Seeing This?".

Appendix

Gobo Interview Script

Introduction

My name is [], and I'm part of a research team working on Gobo. We're conducting interviews to understand how people use different social media platforms and how Gobo might provide an alternative space for managing and exploring social media content. This interview will be no longer than 1 hour. Please let us know if you have any questions or if at any point you want to stop the interview.

Is it okay if we audio record this interview?

Interview Questions

Social media use

- How often do you use social media?
 - When do you use it?
- Which platform do you use most often?
 - What do you mostly use it for? Can you walk me through a typical day?
 - Which features do you engage with the most?
 - What do you like about them? Dislike about them?
- How often do you use your other social media accounts? Do you use them for different purposes?
 - Do you communicate with different people on different platforms? Are the modes of communication different?

- What features do you engage with the most on these platforms?
 - What do you like about them? Dislike about them?
- What is the most recent platform you signed up for? What motivated this?

Frustrations with social media

- When you first started using social media how did you use it? How do you use it today?
 - Are there platforms you don't use as much anymore? Accounts you deleted? If so, why?
 - Did you ever switch social media platforms to replace an existing one you were using (i.e. from Twitter to Mastodon)? If so, why?
 - Do you use platforms for different purposes now compared to before?
 - How does this compare to how your friends' social media use changed (if at all)?
- How do you maintain multiple social media accounts?
 - Are there any challenges in this?
- Tell me about the last time social media made you frustrated.
 - Was it a particular post? A particular feature?
 - How did you respond? Did you feel like you could do anything about it?

Taking control of social media

- How do you think social media companies influence the content you see?
 - Are there moments when you notice it more?
 - How do you feel about this influence?
 - Are you ever surprised by what you see on your social media feed?

- Are there any strategies you use to keep certain kinds of content your social media feed?
 - For example, have you ever hidden a post or “snoozed”/unfollowed someone? If so, why?
- Are there any features you wish social media platforms had? Why?
 - How would you design a platform or feature differently?

Gobo

- What led you to try Gobo?
- Is there any feedback you wanted to share about Gobo?
 - Features that were helpful? Not helpful?
 - Ideas for new features?

Post-interview

Thank you so much for your time. We really appreciate your participation, and if you're interested, we'll be sure to share our paper with you once it's published. We'll also be sending along the Amazon gift card as compensation for your participation this week. Please feel free to email us in the future with any questions, comments, or concerns!

References

- Angwin, Julia, and Hannes Grassegger. "Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children — ProPublica." ProPublica, June 28, 2017. <https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms>.
- Beer, David. "The Social Power of Algorithms." *Information, Communication & Society* 20, no. 1 (2019): 2.
- Browne, Simone. *Dark Matters: On the Surveillance of Blackness*. Duke University Press, 2015.
- Certeau, Michel de. *The Practice of Everyday Life*. Univ of California Press, 2011.
- DeVito, Michael A., Darren Gergle, and Jeremy Birnholtz. "'Algorithms Ruin Everything': #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, 3163–74.
- Eslamimehdiabadi, Motahhare. "Participating and Designing around Algorithmic Socio-Technical Systems." University of Illinois at Urbana-Champaign, 2019.
- Eslami, Motahhare, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. "First I 'like' It, Then I Hide It: Folk Theories of Social Feeds." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2016, 2371–82.
- Florini, Sarah. "Tweets, Tweeps, and Signifyin'." *Television & New Media* 15, no. 3 (2014): 223–37.
- Harris, Brent. "Preparing the Way Forward for Facebook's Oversight Board." About Facebook, January 28, 2020. <https://about.fb.com/news/2020/01/facebooks-oversight-board/>.

- Hirsch, Tad. "Surreptitious Communication Design." *Design Issues* 32, no. 2 (2016): 64–77.
- Hirsch, Tad, Kritzia Merced, Shrikanth Narayanan, Zac E. Imel, and David C. Atkins. "Designing Contestability: Interaction Design, Machine Learning, and Mental Health." *Proceedings of the 2017 Conference on Designing Interactive Systems*, June 2017, 95–99.
- Hogan, Bernie. "From Invisible Algorithms to Interactive Affordances: Data after the Ideology of Machine Learning." *Roles, Trust, and Reputation in Social Media Knowledge 1 Markets, Computational Social Sciences*, November 1, 2013, 2.
- Hosanagar, Kartik, and Vivian Jair. "We Need Transparency in Algorithms, But Too Much Can Backfire." *Harvard Business Review*, July 23, 2018.
<https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire>.
- Hwang, Hyesun, and Kee-Ok Kim. "Social Media as a Tool for Social Movements: The Effect of Social Media Use and Social Capital on Intention to Participate in Social Movements." *International Journal of Consumer Studies* 39, no. 5 (2015): 478–88.
- Know Your Meme. "Wypipo," March 21, 2019. <https://knowyourmeme.com/memes/wypipo>.
- Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, et al. "The Science of Fake News." *Science* 359, no. 6380 (March 9, 2018): 1094–96.
- Pariser, Eli. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin, 2011.
- Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.
- Rader, Emilee, and Rebecca Gray. "Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 2015, 1.

Seaver, Nick. "Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems." *Big Data & Society* 4, no. 2 (2017): 3.

Van Dijck, José. *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press, 2013.

Velkova, Julia, and Anne Kaun. "Algorithmic Resistance: Media Practices and the Politics of Repair." *Information, Communication & Society*, 2019, 1–18.