

/u/thebenshapirobot

Alejandro Ruiz and Julia Witte Zimmerman

May 31, 2022

Introduction

Disinformation, especially from right wing sources, is a commonly accepted fact of everyday internet usage. And yet, even in the third year of a pandemic in which disinformation is responsible for significant loss of life, new methods to combat it remain elusive. Broadly speaking, there are two commonly accepted approaches for dealing with disinformation: journalism and deplatforming. In this case study, we hope to explore an alternative method inspired by systems analysis and the security principles of threat modeling, tailored specifically to potential fans of Ben Shapiro. We think similar processes could be used to combat other sources of disinformation.

Antidisinformation Techniques

Simplifying Assumptions

All models must make simplifying assumptions. For the purposes of this paper, we'll call information that is dangerous and spread knowingly for political (and financial) gain "disinformation". There is much excellent work on

understanding the differences between kinds of speech[1] and the approaches to handling them, but these distinctions are not important here. We will assume that spending time on Ben Shapiro's content, on Reddit, and on YouTube grants us sufficient working understanding of his ideology, his appeal, and his fanbase. Finally, we will assume a model of the internet as a relatively lawless place, governed not by laws but by site admins regrettably beholden only to shareholder value.

The goal of this paper is not to review other work, even cursorily, but to share the results of this case study as quickly as possible, in the hope that this contribution will be helpful to other researchers, information consumers, and activists, even without the context a more formal paper would provide.

Existing Approach - Defense in Depth

Cybersecurity structurally favors the attacker. Bad actors attempting to penetrate a system need only breach it once, whereas defense requires rebuffing every attack in perpetuity. Dealing with instances of disinformation by attempting to debunk them, or waiting for a single piece of disinformation so egregious that the bad actor should be removed from a social media platform, are necessarily actions taken once the disinformation is already in the system.

Defending against attackers that are already in the system is known in the cybersecurity community as defense in depth[2]. Defense in depth is normally implemented as a redundant measure in conjunction with primary strategies, such as network firewalls or password protection. Defense in depth, although expensive and difficult, is worthwhile to intercept the portion of bad faith actors who penetrate past the primary defenses, as the cost of bad actors actually entering the system is considerably greater.

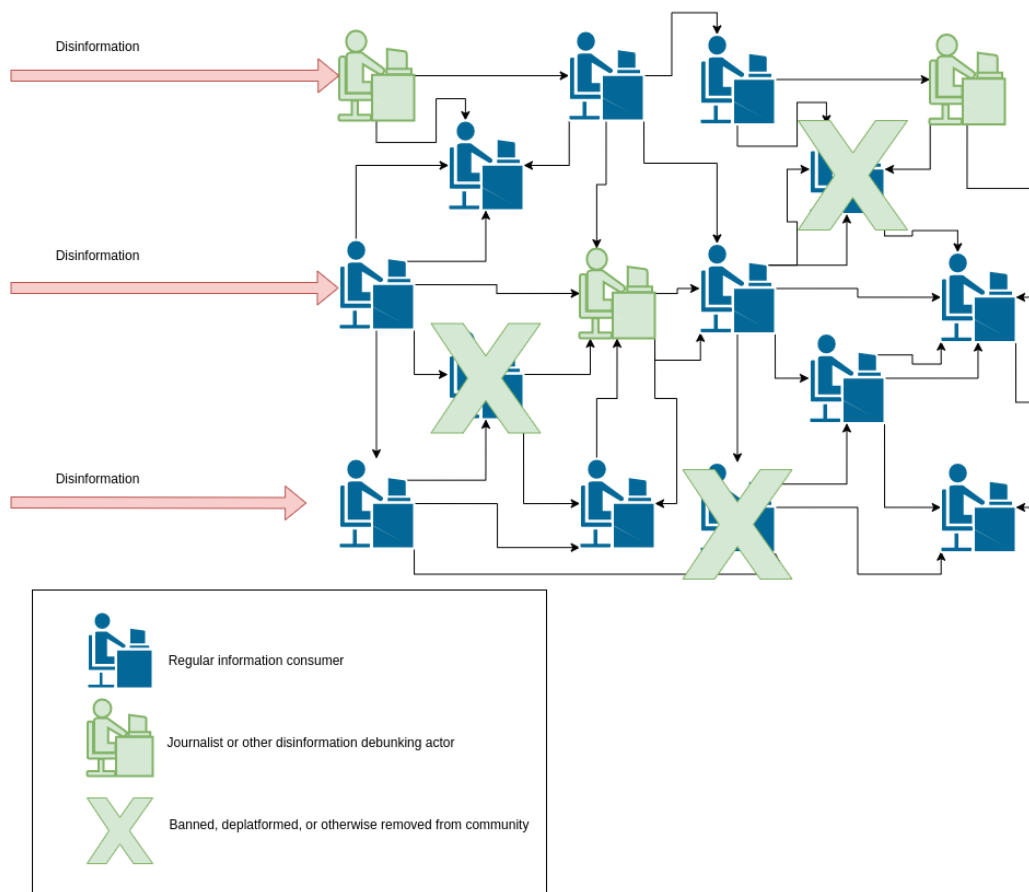


Figure 1: Debunking content or banning users to stop the spread of disinformation within the system is inefficient

Proposed Approach - Adding a Firewall

/u/thebenshapirobot is an attempt to create an antidisinformation firewall by identifying how disinformation enters the system and interrupting that process. Before we can implement our firewall, we need to understand four things: the mechanisms bad actors use to inject disinformation into a system, who Ben Shapiro is, how he launders his content through the mainstream media, and who his fanbase and potential fans are.

Ben Shapiro is the central source in a disinformation network, which disseminates its disinformation primarily via a core fanbase. Although occasionally Ben Shapiro does disseminate content directly through or for mainstream sources, the vast majority of his content - his podcast, his books, his YouTube videos¹ - is aimed at his core fanbase, who then go on to spread his content more broadly.

Rather than targeting people who are already avowed Shapiro fans, we aim to prevent new fans from being created. Specifically, we want primarily to prevent vulnerable people who could become his fans from so converting, so that they do not join the disinformation actors amongst his fanbase. Secondly, we want to counteract his mainstream, respectable image with people who are only tangentially aware of his content, by bringing to their attention some of the plentiful reprehensible (and ludicrous) things he has said and written. In these two ways we aim to create a barrier keeping two kinds of people who could go on to spread Shapiro-originated disinformation for different reasons from disseminating that disinformation any further. Hopefully this firewall can prevent some disinformation promulgation throughout the system.

¹Because of Ben Shapiro's history of weaponizing statistics, citations, and the credibility of other outlets, we have made the editorial decision not to cite directly to his content and we encourage others to do the same.

Ben Shapiro

Ben Shapiro is the editor emeritus of The Daily Wire, the host of the Ben Shapiro show, a frequent speaker for Young America’s Foundation, and consistently in the daily top 10 performing authors of content on facebook [3][4]. He is a hugely influential figure whose content features, as we will show, climate denial, transphobia, racism, calls for ethnic cleansing, and COVID disinformation.

Laundering His Image

Using Mainstream Figures

In order to discuss how Ben Shapiro’s image is laundered in the mainstream, we must, unfortunately, give a few examples. Without this context, readers may find they hold a more charitable view of Ben Shapiro than is warranted due to the very phenomenon we are discussing. It is out of scope for this paper to have a full discussion on Ben Shapiro’s views, so we will limit ourselves to a handful of excerpts from a single speech he gave at the University of Missouri with almost four million views on YouTube[5]:

- “Heterosexual marriage is the cornerstone of society; homosexual marriage offers no benefits to society.”
- “When it comes to global warming, there are two issues: is there such a thing as the greenhouse gas effect? The answer is yes. Is that something that is going to dramatically reshape our world? There is no evidence to show that it will. Is that something that we can stop? There is no evidence to show that we can.”
- “Even climatologists can’t predict 10 years from now. They can’t explain why there has been no warming over the last 15 years. There has been a static trend with regard to temperature for 15 years.”

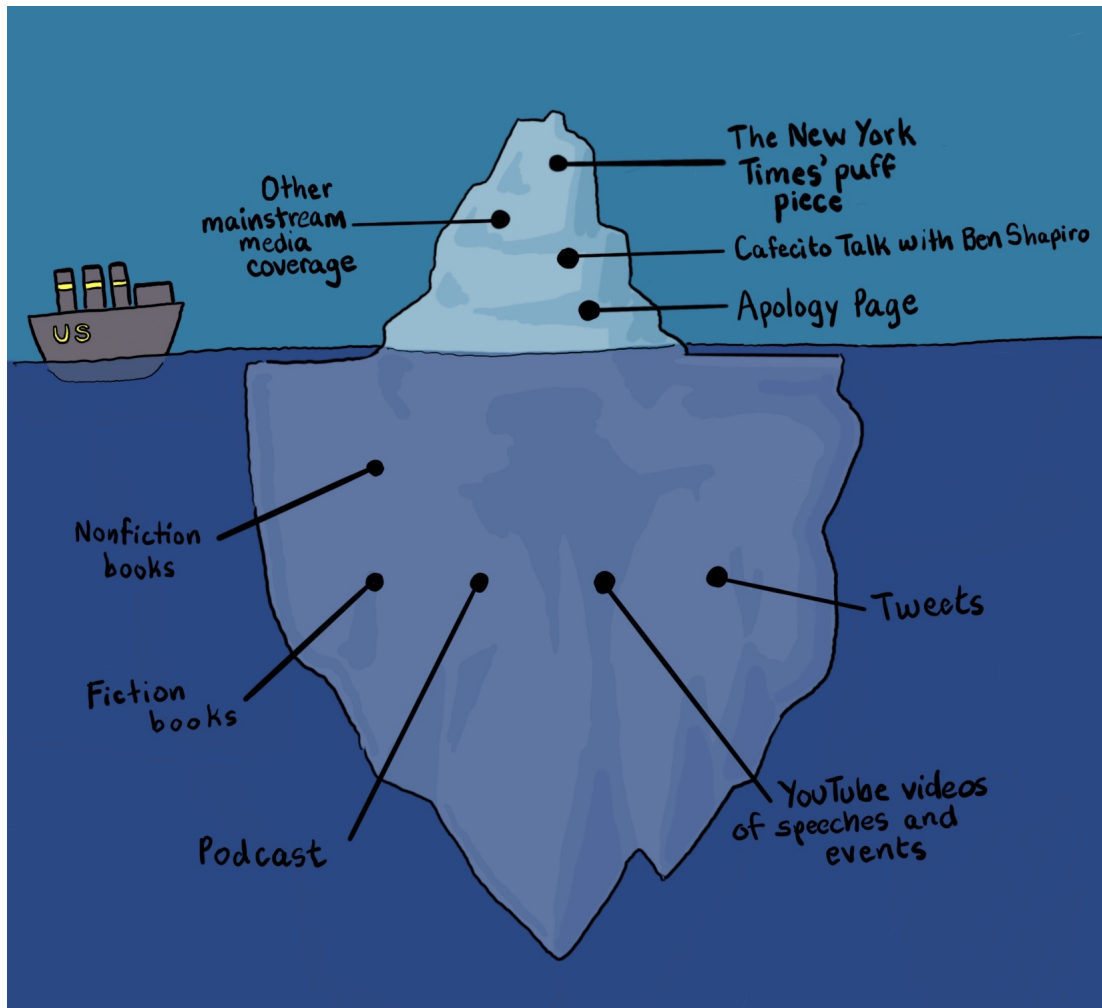


Figure 2: Ben Shapiro carefully manages his image through the curation of layers of content targeted at different audiences. He strategically draws from two classes of content: the underwater, unsavory content that would be rejected by most people, and the above water content that is largely inoffensive.

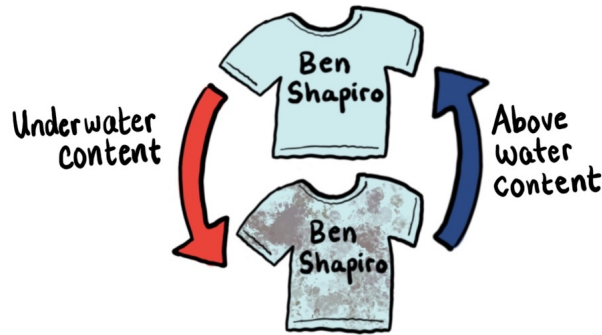


Figure 3: By appearing in mainstream media, Ben Shapiro bolsters his image; by bolstering his image, he makes himself more palatable to the mainstream. He exploits this cycle to disseminate disinformation and hate for fame and profit.

- “As the west grows, and the west’s power grows, the world gets better, and as the west’s power diminishes the world gets to be a worse place.”
- “The idea is that blacks in Americas are more impoverished than whites because it is systemic racism. Okay, according to the Brookings Institute, which is a very left institute, the Brookings Institute has found that if you want to not be permanently poor in the United States it’s actually really easy—it’s a wonderful country. If you don’t want to be permanently poor in the United States you need to do three things: finish high school, get a job, don’t get pregnant before you get married. That’s it. Seriously, if you do those three things, you will not be permanently poor in the United States. Would you like to know why there’s a disproportionate poverty rate in the black community? Because there is a disproportionate single motherhood rate and drop out rate in the black community. As much as we can talk about white privilege, the fact is that the single motherhood rate in the black community in 1960 was 20%, today it is upwards of 70%.”
- “To mature people, the subjective feelings of others are only important if you’re married to them.”
- “The strongest weapon on the left is always character attacks. You’re a racist sexist bigot homophobe. It’s actually the only argument they

have, generally speaking. They disagree with you, therefore you're a bad person. Basically, Barack Obama said this with regard to anybody who objects to his Syrian refugee idea, right? You're a bad person; you hate children; you hate women; you fear widows; you're a bad person. You can't actually have a logical discussion about the issue. You're just a bad person. That's the strongest argument on the left's side. The strongest argument on the right's side is evidence and logic."

Compare these statements to the way he is portrayed by The New York Times two years later, in a piece titled 'Ben Shapiro, a Provocative 'Gladiator,' Battles to Win Young Conservatives.' [6]

- "Mr. Shapiro, 33, a graduate of Harvard Law School, is the cool kid's philosopher, dissecting arguments with a lawyer's skill and references to Aristotle."
- "Liberals may not like his conclusions, but they are guiding young people at a time when the conservative movement is adrift and ideas of white nationalism are competing for their attention. Mr. Shapiro, an Orthodox Jew, was one of the first to call out the alt-right movement, denouncing it as racist and anti-Semitic at a time when most people saw it as counterculture and cool."
- "His aggressive tone draws in audiences, he said, but he does not attack unfairly, stoke anger for the sake of it, or mischaracterize his opponents' positions. He even hits his own side, as he did with Sean Hannity for not weighing in on Roy S. Moore, the embattled Alabama Republican, and Mr. Bannon for supporting him."
- "He takes apart arguments in ways that makes the conservative conclusion seem utterly logical, like putting a key in a locked door."
- "Mr. Shapiro has always been deeply conservative and does not pretend to be objective. But he says his market niche is giving clear-eyed reads of current events, not purely partisan rants."

The piece makes no mention of his climate change denialism. In his speech, he blames the black community for their own poverty, yet the Times portrays

him as a conservative alternative to white nationalism, saying “Liberals may not like his conclusions, but they are guiding young people at a time when the conservative movement is adrift and ideas of white nationalism are competing for their attention. Mr. Shapiro, an Orthodox Jew, was one of the first to call out the alt-right movement, denouncing it as racist and anti-Semitic at a time when most people saw it as counterculture and cool.”[6] The piece lets the description that Ben Shapiro provides “clear-eyed reads of current events, not purely partisan rants” go unchallenged. This is an irresponsible puff piece uncritically repeating what Ben Shapiro wants his mainstream brand to be.

Unfortunately, the Times is not alone. Credible outlets and figures continuously lend their credibility to Ben Shapiro, including the BBC[7], CNN[8], PBS[9], Politico[10], Newsweek[11], the Mayor of Miami[12], and local newspapers via syndicated columns[13].

Apology Page Strategy

Ben Shapiro uses media outlets’ credibility to bolster his own. What sets him apart from aspiring journalists and commentators is that he uses them in bad faith; he uses his inclusion in other sources as a means of laundering his own content.

As a particularly direct example of how he launders content, take his apology page[14]. Sometimes, something he says is so offensive that it causes viral outrage. To protect his image from the backlash, he has created a page on his site titled “So, Here’s A Giant List Of All The Dumb Stuff I’ve Ever Done (Don’t Worry, I’ll Keep Updating It).”

When an interviewer calls him out on one of these things he’s said, he’ll defuse the attack by claiming it’s on his apology page. He deploys this knowing full well that no one reads past the headline, because the page is full of bad faith apologies, non-apologies, and, in some cases, outright refusals to apologize.

Take, for one of many examples of how inadequate his apologies are, his famous “Enemy Civilian Casualties” column [15], which contains the following

quotes, and its subsequent entry on the apology page:

- “I am getting really sick of people who whine about “civilian casualties.” Maybe I’m a hard-hearted guy, but when I see in the newspapers that civilians in Afghanistan or the West Bank were killed by American or Israeli troops, I don’t really care. In fact, I would rather that the good guys use the Air Force to kill the bad guys, even if that means some civilians get killed along the way.”
- “It doesn’t matter to me if some of their “civilians” get killed for involvement with the enemy. I’m glad the U.S. military decided to use a massive air campaign rather than going in full force with ground troops.”

The thesis of this article is that it is acceptable to kill civilians, especially non-American, non-white ones, who are essentially the “bad guys”. Here is the full text of the apology he made for it:

“In this column, written when I was 18, I suggested that civilian casualties in war were of no concern. While the larger point of the piece — that we must calculate the risk to American service members when we design rules of engagement — is partially correct, the piece is expressed in the worst possible way, and simplifies the issue beyond the bounds of morality (particularly by doubting the civilian status of some civilians). It’s just a bad piece, plain and simple, and something I wish I’d never written. It’s also good evidence that a lot of the stuff you think is smart at 18 is just you being an idiot at 18.”

Although parts of this apology might superficially sound good, it mischaracterizes the offensive aspects of his initial work, and even defends some aspects of it. But Ben Shapiro nevertheless points to these apologies when called out, deflecting criticism in the moment and using his apology page to rehabilitate and launder his reputation and content. He can now distance himself from some specific piece of his own content without actually repudiating it, and

potentially while continuing to spread the same slightly repackaged ideas (especially in content aimed at his core fanbase, who are unlikely to take offense).

The Appeal of Ben Shapiro

In order to understand what content might be effective for our bot to deploy, we need an understanding of Ben Shapiro fans and his appeal. We used our own experiences with his content and fandom as well as sociolinguistic analysis to understand what kind of person might find Ben Shapiro appealing. The Ben Shapiro bot is predicated on this understanding of his fans and potential fans, and aims to present Ben Shapiro’s content to vulnerable, potential fans in a way that makes maintaining such an identity difficult and discourages fandom.

Facts don’t care about your feelings

Ben Shapiro sums up his ideology and his appeal with his catchphrase – immortalized in the titles of two of his books² – “facts don’t care about your feelings.” The phrase privileges facts over feelings, something scientific and rational as well as stereotypically masculine. Drawn to this ideology, Ben Shapiro fans use linguistic cues to shore up their identities as rational, with undertones of whiteness, masculinity, and non-conformity. They do this in a way similar to the way young men make use of the phrase “dude” to construct camaraderie, heterosexuality, and non-conformity[16], and nerds make use of superstandard English to construct intelligence and non-conformity[17].

²Because of Ben Shapiro’s history of weaponizing statistics, citations, and the credibility of other outlets, we have made the editorial decision not to cite directly to his content and we encourage others to do the same.



Figure 4: Usage of the 3-gram “facts don’t care” spikes in 2019, reflecting the spread of Ben Shapiro’s catchphrase through popular culture.[18]

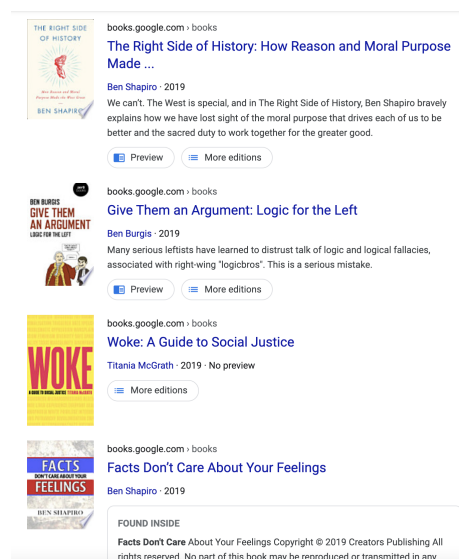


Figure 5: In the Google books corpus, the 3-gram “facts don’t care” was often used in the same vein as, or in direct reference to, Ben Shapiro. He also used the phrase within his own books.[18]

YouTube

YouTube is often cited as the platform on which fans discovered Ben. This makes sense, because Ben Shapiro’s YouTube presence is carefully tailored to his intended audience. Consider the clickbait titles of some of his YouTube videos:

- Ben Shapiro DESTROYS Transgenderism And Pro-Abortion Arguments
- Ben Shapiro Destroys Feminism
- Ben Shapiro SMACKS DOWN Black Lives Matter

These titles may seem absurd, but they are effective. They are aimed primarily at teenage boys and young men, targeting them at a time in their life when they are first grappling with the complexities of patriarchy, systemic racism, and their own role in these large scale societal problems. These new recruits are not necessarily hateful; they are vulnerable, and what they find in Ben Shapiro is a relief. He makes the world simple and black-and-white. He assures them that good people are good, bad people are bad, men should be men, and women should be women. He asserts America is the best country. He “understands” that Muslim people and black people and trans people are different and therefore scary. Hearing all this from a respected authority figure is an incredible social relief from the difficulties of transitioning into adulthood. He provides clarity and simple rules for understanding a complex world. Ironically, despite this siren call, the image he builds for himself and his fans is one of rational truth-seeking and free-thinking.

Ben Shapiro’s fans must negotiate their identities as fans of his repugnant content with their identities of themselves as moral, logical people – an identity Ben Shapiro often reinforces, with quotes like, “the strongest argument on the right’s side is evidence and logic.”.[5] This requires holding in balance two sides of Ben Shapiro’s rational identity – the respectable and the non-conformist. By allying his most offensive content with nonconformity, fans can explain its lack of acceptance by the mainstream while maintaining that they are moral and rational. This means that nonconformity, free-thinking,

and rationality are all key to the identity of Ben Shapiro fans once they're Ben Shapiro fans – but it also means that people with such an identity may be more apt to find him appealing, especially if they are vulnerable for other reasons, such as being young and wanting to define themselves in a complex world, or feeling attacked by changes to the status quo (like increasing LGBTQIA visibility and acceptance, the Black Lives Matter movement, or feminism, to name just a few).

Ben Shapiro's fans in their own words

We looked at comments in the Ben Shapiro subreddit, a subsite of reddit with more than 52,000 users. By looking at comments within a subreddit specific to Ben Shapiro, we allowed people to self-identify as likely fans of Shapiro (some people profess to be on the subreddit ironically[19]). We also looked at YouTube videos made by fans/ former fans about Ben Shapiro.

Of course, people's identities are not defined solely by their online comments, their video content, or how they speak, but we think it's reasonable to make use of fan-created comments and video content to get insight into the people subject to Ben Shapiro's appeal.

Although Ben Shapiro's fans don't necessarily spurn coolness, they do share with nerds “a particular emphasis on language as a resource for the production of an intelligent and nonconformist identity.”[17] For example, in this comment,

“I love when liberals are like “he said THIS” and proceed to provide a sensible and valid take on something”[20]

the use of the phrase “sensible and valid”, and in particular the use of “valid” from scientific and mathematical discourse, reinforces the rationality of the speaker (and of other Shapiro fans, in comparison to “liberals” who can't understand).

We found that Ben Shapiro's fans often make use of linguistic strategies of

register and lexical choice to “invok[e] discourse genres and topics associated with intelligence.”[17] They do this because they want to use the process of iconization to establish their rational identities. “In iconization, the ideological tie between form and social meaning is stronger still: the characteristics of a language are seen as a reflection of the essential characteristics of its users. The relationship is therefore not merely one of (perceived) juxtaposition, as in indexicality, but of (perceived) resemblance.”[17]

Ben Shapiro subreddit

To find specific evidence of this practice, we went to the top most up-voted post of all time in the subreddit[21], and started reading the comments in order, that way we would be looking systematically through content that is fair to take as representative of the community as a whole.

Within those comments we found this exchange,

“I agree with you. The Democrats are guilty of projection through accusations. They claim Republicans are doing exactly what they are trying to do. I think the Democratic Party has some seriously mentally ill people in charge.”[22]

“Sadly the real problem is that around 75% of the population has a IQ of 110 or less. They are evenly distributed on both sides... fools listening to fools... mob rules.”[23]

“Projection through accusations” has an air of clinical diagnosis, as does “mentally ill”, and in the second comment, the lack of intelligence of the general population (excluding the people in the conversation) is explicitly called out as the root of other political problems.

Finally from within that thread, we look at,

“I didn’t know those 3 points were actually ever contested or argued against... vaccine doesn’t prevent transmission, the vulnerable are more susceptible, and death from is not the same as death with... It’s insane how all that’s not common sense; How do we evolve from this point as a species?” [24]

Using long words like “contested”, “transmission”, “vulnerable”, and “susceptible” emphasize the speaker’s large vocabulary, and thus education – as does throwing in a semi-colon. Calling out explicitly that these things should be common-sense (something Ben Shapiro often does) once again positions this speaker as intellectually superior to the masses.

Additional fan content

We found a video where a self-professed Ben Shapiro fan, who is not a native English speaker, calls in to ask about who he could watch debate on the political left. In the preamble to the question, he talks about his “fascination” with the concepts Ben Shapiro covers and the “research” he did to try to find the kind of videos he was looking for. Framing his search as research elevates it from entertainment and helps build the caller’s rational identity. The caller also references that Ben Shapiro’s ideas are billed as “common-sense”. [25]

In another video, a former fan (who is now less conservative) reflects on what drew her to Ben Shapiro, Jordan Peterson, and other right-wing commentators, and describes it as a “truth-seeking narrative” which is “a tactic that these types of right-wing commentators often use to lure people in”. [26] She also says that Ben Shapiro’s readiness with statistics and facts at hand during debates was alluring, and added to her feeling that the truth had been obscured by the mainstream media and her left-leaning education – that these forces were trying to control her. She says, tying her identity to non-conformity,

“If you know me, I like being an independent thinker. I like feeling as though I have control over my thoughts. I’m someone who cares about finding what truth is – or, I used to be. Now I

am not quite certain that truth exists, but that's a whole 'nother topic."

and,

"I think it almost comes from that instinct of when I was younger, of rooting for the underdog, but now it was directed towards the underdogs of thinking, if you could call it that – people like Ben Shapiro is not what you see in your instagram infographics. Ben Shapiro is not what you hear in your daily news; it's not who you learn about in school." [26]

Note that besides "nother", very few informal forms are used, despite the informal, unscripted format of the video. "I'm" is used once, as is "it's". The other words are used in their full forms, as was often seen in the "super-standard" speech of nerds. "Superstandard English contrasts linguistically with Standard English in its greater use of "supercorrect" linguistic variables: lexical formality, carefully articulated phonological forms, and prescriptively standard grammar. It may also go beyond traditional norms of prescriptive correctness, to the point of occasionally over-applying prescriptive rules and producing hypercorrect forms." [17] In superstandard English as well as in the language of Ben Shapiro's fans, marking the speaker's intelligence is of utmost priority, although the deviations from standard English may be small: "By the same token, the superstandard need not deviate substantively from the colloquial standard in order to be considered distinctive; because it is marked with respect to Standard English forms, even relatively slight use of supercorrection and hypercorrection can call attention to itself. Superstandard English is therefore a marked variety that may contrast ideologically both with the unmarked colloquial standard and with marked nonstandard English. However, because it draws on the prescriptive standard, it also contributes to the linguistic ideologies that elevate one linguistic variety over others." [17] The language of Ben Shapiro's fans hearkens back to whiteness and masculinity in part through its similarity to the language of nerds. [17]

Finally, another former fan (who is now more conspiratorially-minded and more conservative) says of Shapiro, "he's basically a human version of Google" [27].

This comparison makes explicit the rational persona of Ben Shapiro – he is equivalent to a computer – and the speaker attributes her respect for Ben Shapiro to it.

Analysis of reddit comments

In order to complement the few examples of fan content we delved into in detail previously, we scraped all the comments from the top 100 posts for the week of April 24th³ from several subreddits: /r/benshapiro, /r/politics (a message board for discussion of politics), and /r/all (an aggregator of popular posts from other message boards). This yielded 1310 comments for /r/benshapiro, 13611 for /r/all, and 11993 comments for /r/politics. We converted these comments into vectors of 1-grams. We then compared these corpuses to each other via the wordshift package[28] (the wordshifts are fairly large diagrams, so please see the following pages). We found that the words “wrong”, “argument”, and “against” impacted the happiness score relatively more in the comments of /r/benshapiro than either /r/all or /r/politics, one interpretation of which is that /r/benshapiro has a more empirical tone. The overall happiness score, based on the labMT lexicon[29], for the words used in /r/benshapiro was 5.66.

Certain topics – as indicated by individual words – were more likely to come up in /r/benshapiro than /r/all or /r/politics, including “black”, “gay”, and “Biden”. The fact that “black” and “gay” appear relatively more often in /r/benshapiro could be interpreted as evidence of the importance of whiteness and heterosexuality to the community, and therefore to the individuals constructing their identities within that community.

³Retrieved on April 25, 2022

Biden	72
get	70
would	69
no	69
will	68
right	68
This	64
him	64
his	63
has	63
more	62
at	61
by	59
my	56
They	55
out	52
don't	51
don't	50
how	49
know	48
even	48
same	46
see	46
Trump	45
Ben	45
me	44
should	44
then	43
way	43
over	43
been	43
being	42
were	42
want	41
black	41

This table includes some of the more frequent 1-grams used in the comments

from /r/benshapiro. Note that “black”, “him”, and “his” are relatively often used, potentially reflecting the importance of whiteness and masculinity to the community.

Cringe as an antidote to rationality

Ben Shapiro’s fans use linguistic cues including register choice and lexical choice to construct a non-conformist and intelligent identity, which we’re summing up as rational. This identity includes aspects of whiteness and masculinity, both in the content talked about, and in the manner of speech (including by similarity to the speech of nerds[17]). A rational identity is especially valuable for Ben Shapiro’s fans, many of whom are young men, “as they navigate cultural Discourses of young masculinity, which simultaneously demand masculine solidarity, strict heterosexuality, and nonconformity.”[16] Ben Shapiro fans use their rational identities to balance two important and opposed forces: respectability (for their arguments and conclusions) and non-conformity (for their free-thinking).

We think that inducing a feeling of “cringe” contradicts this rational identity, and is therefore a way to provoke a crisis of identity which his potential fans may resolve by distancing themselves from Ben Shapiro.

We don’t think fact-checking Ben Shapiro’s claims will necessarily have this effect, since offering a pair of opposing facts gives the appearance of two equally-likely options to choose between. Facts can be argued, and Ben Shapiro is skilled at rhetorical strategies that make him look as if he is well-armed with facts. A potential fan of Shapiro may see Shapiro’s point, especially when presented as Shapiro wants it presented to that person at that time, and fail to be swayed by a subsequent rebuttal. Furthermore, a reddit bot can’t engage (yet) in a detailed argument for long, so that strategy is implausible. We need concise, tailored, yet pre-packaged content for our bot given the technical limitations, so we need a strategy that will work with that content. Finally, facts are not identity – someone can be shown to be wrong about a fact without that altering your overall perception of them, whereas we want to inoculate potential fans wholesale against Ben Shapiro’s appeal.

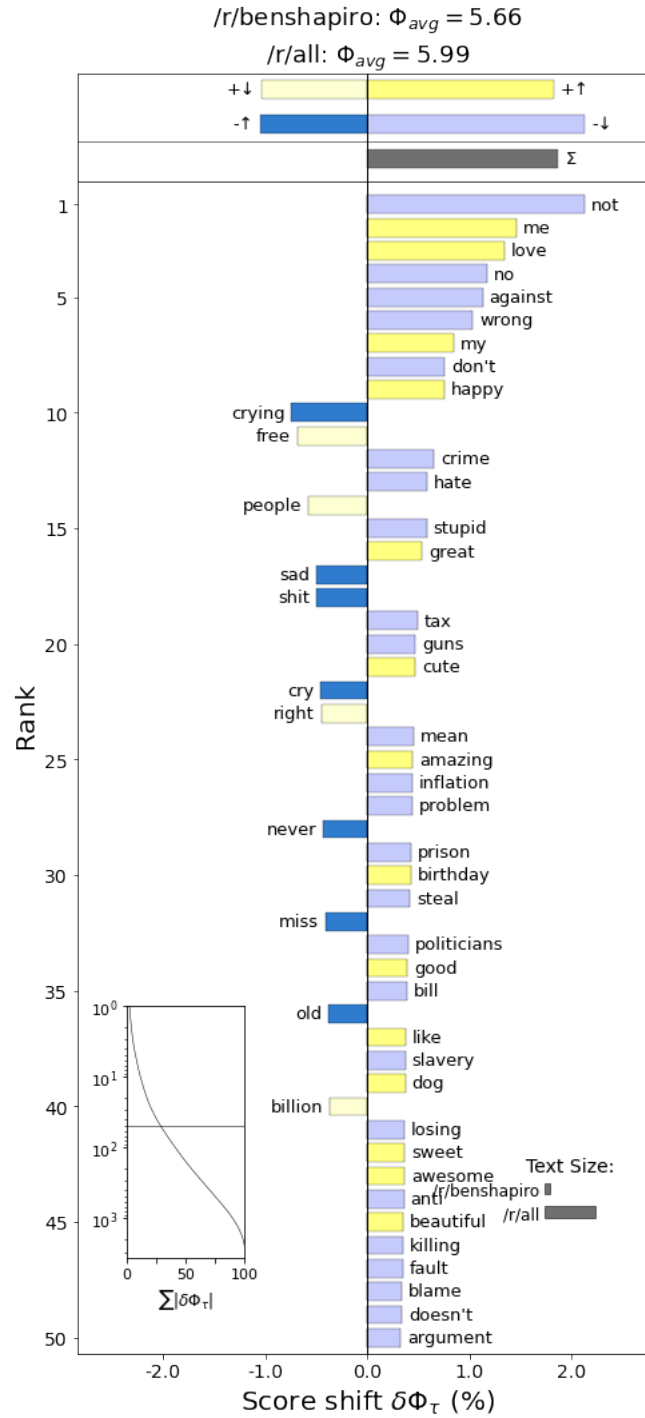


Figure 6: The overall happiness score for the words used in /r/all was 5.99.

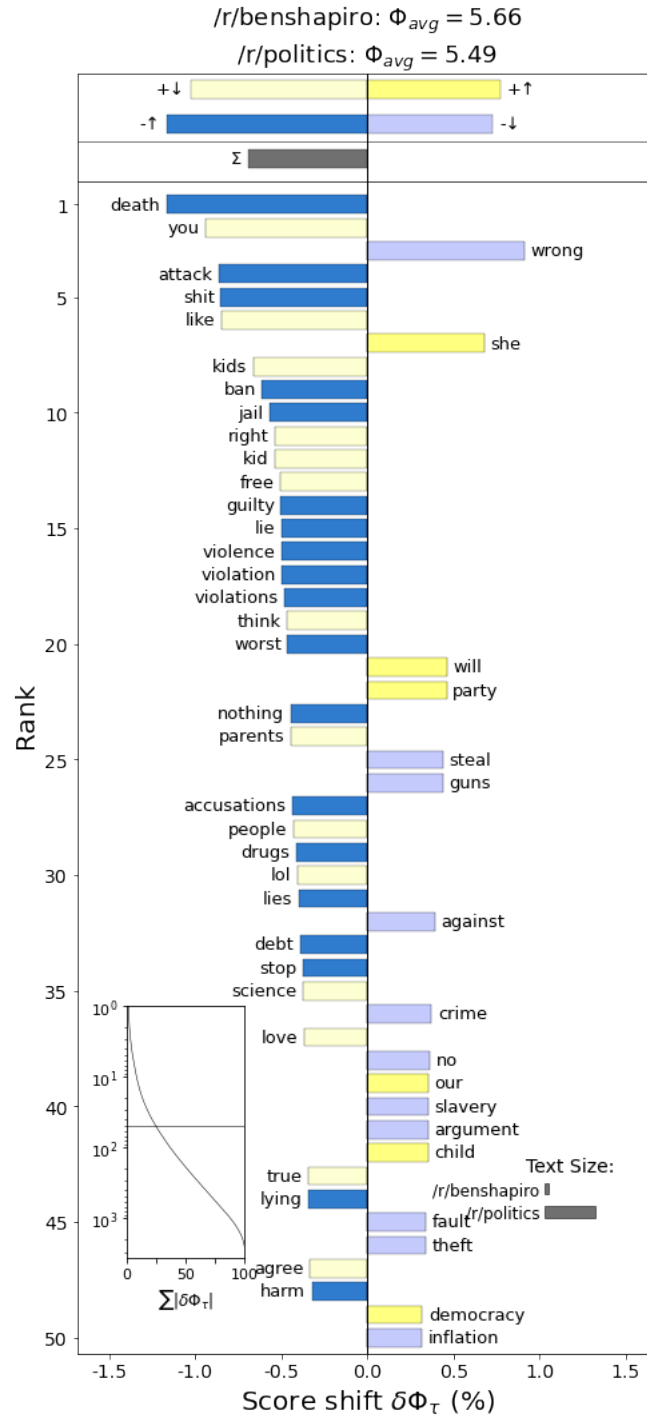


Figure 7: The overall happiness score for the words used in /r/politics was 5.49.

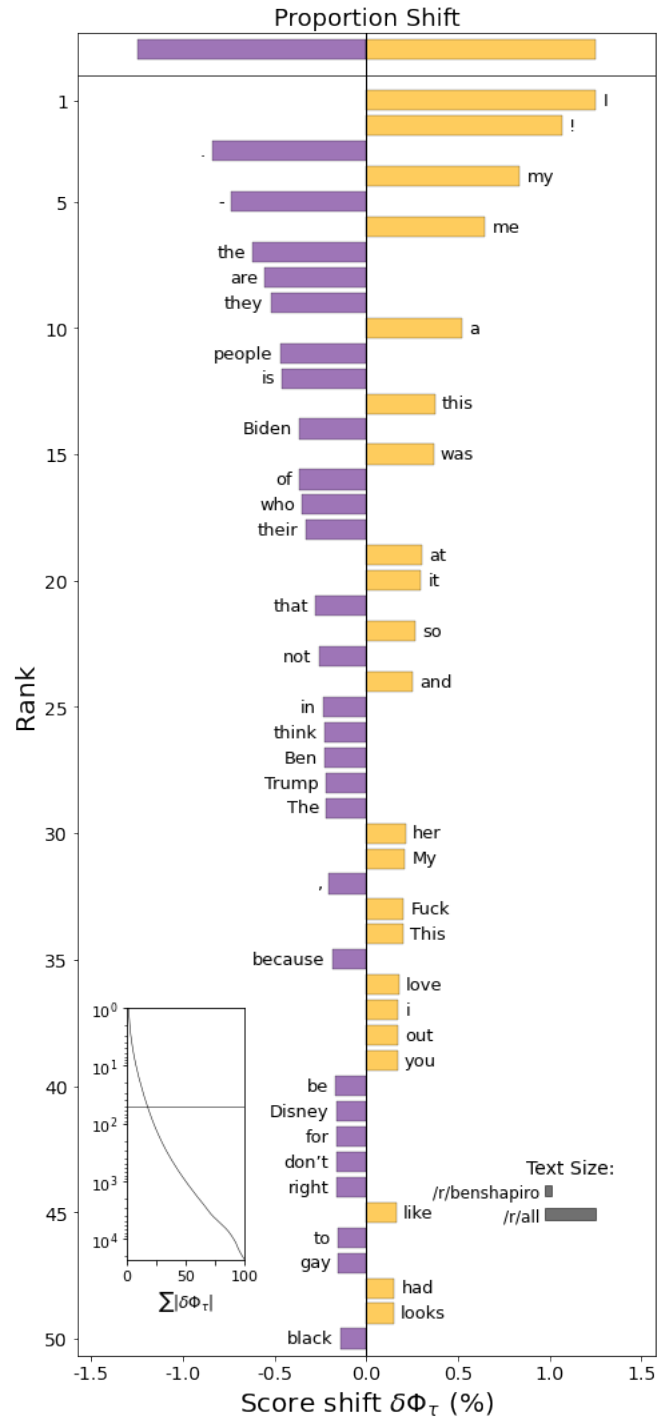


Figure 8: Certain words were more likely to occur in /r/benshapiro than /r/all.

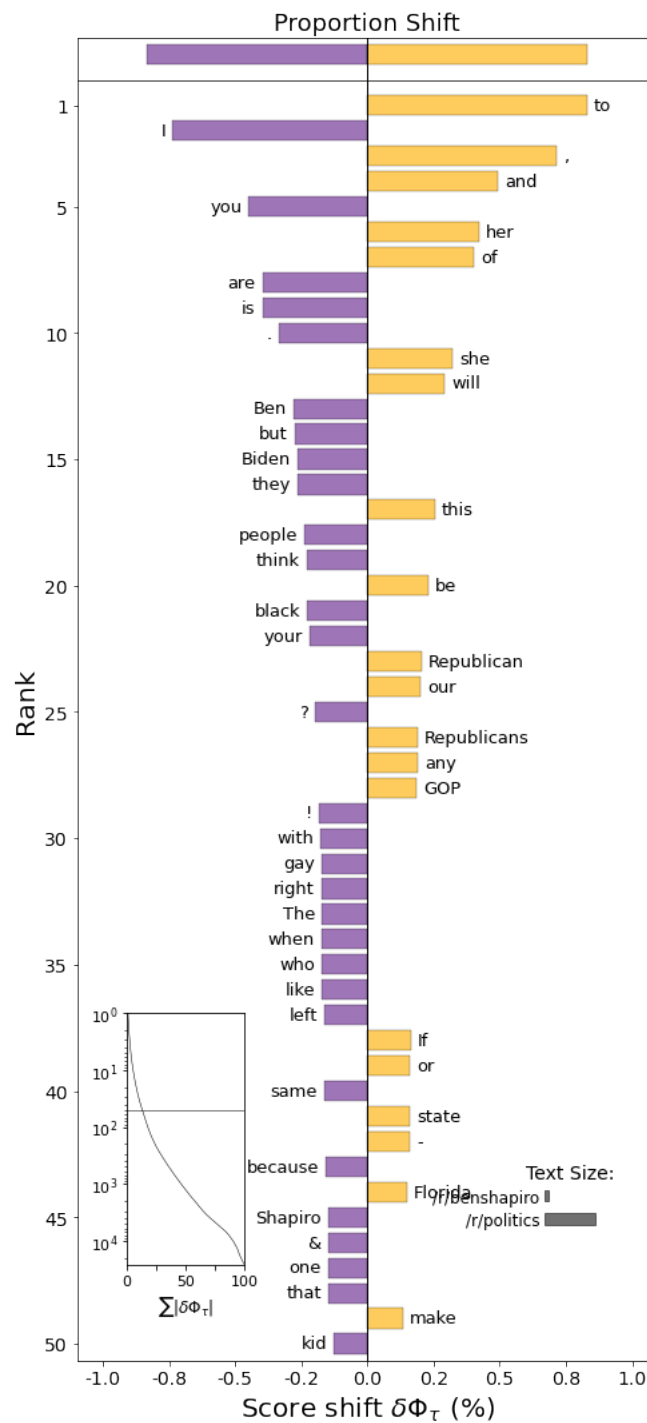


Figure 9: Certain words were more likely to occur in /r/benshapiro than /r/politics.

This is how we came to our strategy of inducing “cringe”. Presenting someone as “cringe” is antithetical to their being someone you could plausibly take seriously. It is the opposite of respect. Having established the importance of rationality, non-conformity, and free-thinking to his fanbase, we think those qualities are likely to be important to his potential fans as well, and given that, if they cannot respect Ben Shapiro, we think they are unlikely to want to further integrate him into their belief system.

We aim to induce “cringe” in the following ways:

- By directly quoting some of Ben Shapiro’s most offensive content
- By directly quoting some of Ben Shapiro’s most ridiculous content
- By equipping the bot with a sense of humor (lack of self-awareness and lack of humor are highly correlated with “cringe”)
- By inspiring other users to comment derisively about Ben Shapiro, potentially making him the butt of jokes and exposing vulnerable people to their peers’ low opinion of Ben Shapiro
- By making a user arguing with the bot look “cringe” themselves

We will go more into the technical implementation of these features in the *Bot Behavior* section.

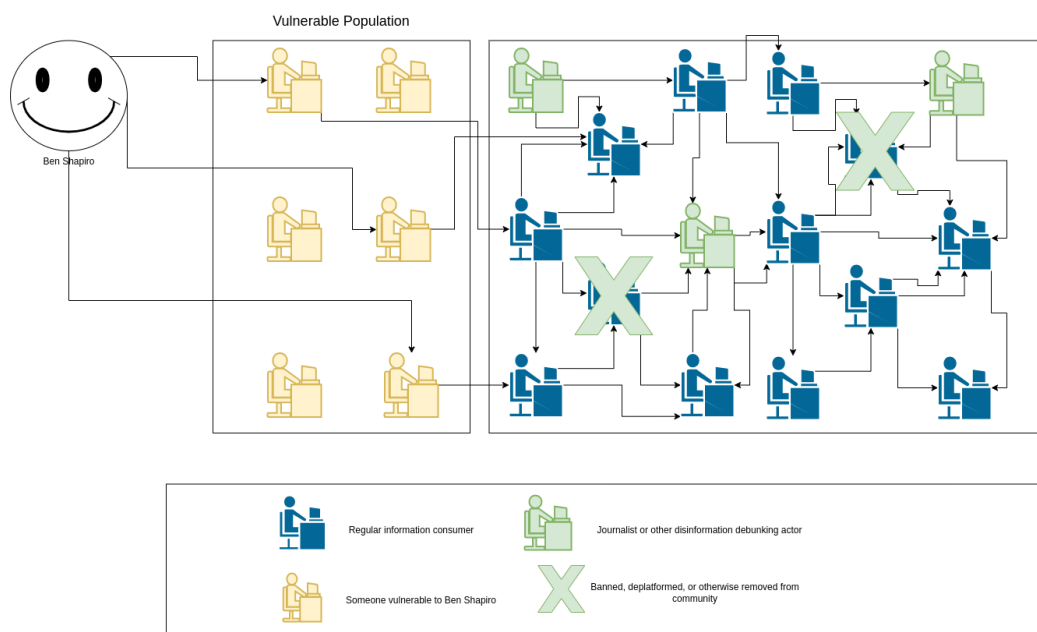


Figure 10: This diagram shows how Ben Shapiro targets a vulnerable population of mostly young, naive men. From there, his ideas metastasize. His laundered image allows his targets (and others) to take him seriously when they see him on YouTube or elsewhere.

Ben Shapiro's Disinformation Dissemination System

`/u/thebenshapirobot`

Goals

The bot's primary goal is to inoculate users vulnerable to Ben Shapiro by exposing them to Ben Shapiro's "underwater" content and to their peers' low opinion of Ben Shapiro. If the weaponization of recommendations algorithms on various platforms such as YouTube can drive people to him, then we hope the bot can prime Ben Shapiro's targets before that first recommendation by showing that Ben Shapiro is offensive, shameful and embarrassing in their peer groups – before the videos are autoplaying in the background. The bot tries to make Ben Shapiro "cringe" before they start consuming his content, not after. By targeting the exploitable population more directly, we hope to affect his dissemination mechanism further upstream.

Bot Behavior

The bot's core functionality is to respond to every single reddit comment that contains the string 'Ben Shapiro' with a comment in this format:

I saw that you mentioned Ben Shapiro. In case some of you don't know, Ben Shapiro is a grifter and a hack. If you find anything he's said compelling, you should keep in mind he also says things like this:

"If you believe that the Jewish state has a right to exist, then you must allow Israel to transfer the Palestinians and the Israeli-Arabs from Judea, Samaria, Gaza and Israel proper. It's an ugly solution, but it is the only solution... It's time to stop being squeamish."

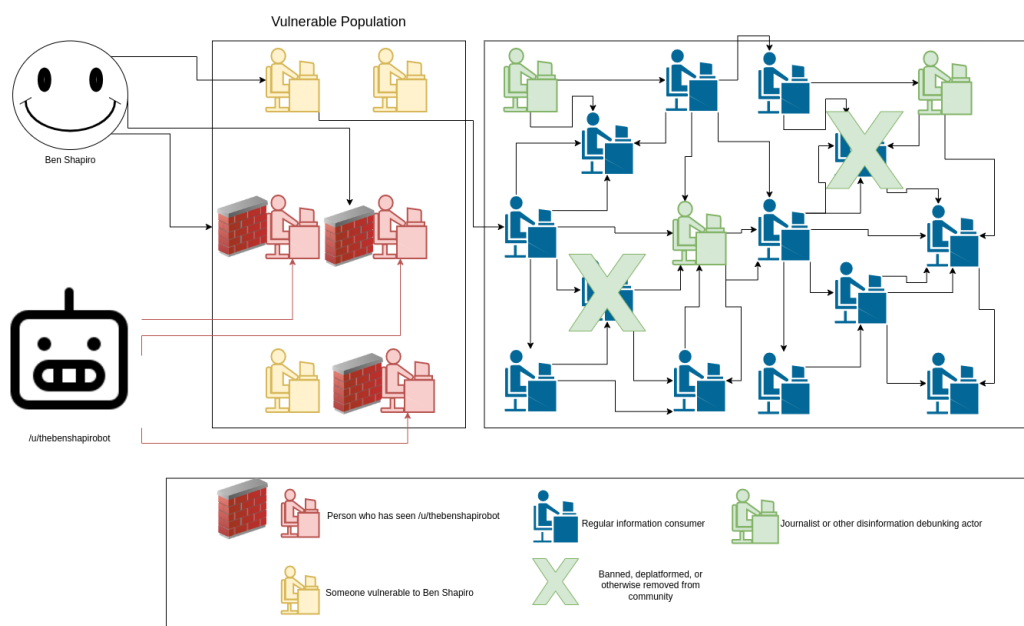


Figure 11: /u/thebenshirobot is designed to act as a firewall for Ben Shapiro's content by inoculating the vulnerable population.

By invoking a quote that proposes ethnic cleansing (as in this example) or other repulsive content, we hope to show users the ugly, unlauded side of Ben Shapiro that actually comprises most of his content.

For this initial interaction, the bot has a variety of quotes, ranging in seriousness and subject matter. Here are a few example quotes:

- “The Palestinian Arab population is rotten to the core.”
- “Since nobody seems willing to state the obvious due to cultural sensitivity... I’ll say it: rap isn’t music.”
- “If you wear your pants below your butt, don’t bend the brim of your cap, and have an EBT card, 0% chance you will ever be a success in life.”
- “Pegging, of course, is an obscure sexual practice in which women perform the more aggressive sexual act on men.”

These initial interactions generally contain his thoughts on more serious topics, with the occasional juvenile one. Because initial comments are more likely to be seen by a wide audience, these comments have two main goals: de-launders Ben Shapiro’s brand by exposing the general population to the horrific things he has consistently said, and co-opt any conversation about Ben Shapiro into one guided by the bot, which will hopefully lead to Ben Shapiro being portrayed unflatteringly in one of the ways discussed in the *Cringe as an antidote to rationality* section.

Once a conversation has been co-opted by the bot, it has follow-up behaviors designed for the to seed situations that make Ben Shapiro and his fans look cringe. The bot will parody Ben Shapiro by mimicking his mannerisms, for example, by challenging people to debates and never letting people have the last word. If someone tells the bad “bad bot,”⁴ it will respond “so much

⁴A Reddit user has created a system for tracking the community’s response to bots; when users comment good or bad bot, those comments are tallied here: <https://botrank.pastimes.eu/>.

for the tolerant left,” or other similarly sarcastic things. It will also quote particularly offensive or stupid things he has said, depending on context.

Comments further down the thread are less likely to be seen by passers-by, so less emphasis is given to de-laundersing his content. Instead, the follow-up quotes are more wide ranging and absurd, including everything from excerpts from his obscure 2016 novel *True Allegiance* to his famous WAP tweet.⁵

The bot will also look for keywords and reply with a relevant thing Ben Shapiro has said in these follow up comments. Keyword topics range from “climate” “healthcare,” or “covid,” to “sex” and “novel.” Most of the time, the relevant things are simply quotes.

Some examples:

- “Let’s say your life depended on the following choice today: you must obtain either an affordable chair or an affordable X-ray. Which would you choose to obtain? Obviously, you’d choose the chair. That’s because there are many types of chair, produced by scores of different companies and widely distributed. You could buy a \$15 folding chair or a \$1,000 antique without the slightest difficulty. By contrast, to obtain an X-ray you’d have to work with your insurance company, wait for an appointment, and then haggle over price. Why? Because the medical market is far more regulated — thanks to the widespread perception that health care is a “right” — than the chair market.*”

“*Does that sound soulless? True soullessness is depriving people of the choices they require because you’re more interested in patting yourself on the back by inventing rights than by incentivizing the creation of goods and services. In health care, we could use a lot less virtue signaling and a lot less government. Or we could just read Senator Sanders’s tweets while we wait in line for a government-sponsored surgery — dying, presumably, in a decrepit chair.”

- “Most Americans when they look around at their lives, they think:

⁵Because of Ben Shapiro’s history of weaponizing statistics, citations, and the credibility of other outlets, we have made the editorial decision not to cite directly to his content and we encourage others to do the same.

I’m not a racist, nobody I know is a racist, I wouldn’t hang out with a racist, I don’t like doing business with racists—so, where is all the racism in American society?”

- “I don’t think the law has any role whatsoever in banning race-based discrimination by private actors.”

The “novel” keyword is a particular edge case. Ben Shapiro is a prolific writer. Though he mostly writes nonfiction propaganda, he has occasionally dabbled in fiction propaganda. In 2016, Ben Shapiro wrote a novel titled *True Allegiance*. The book deals with a wide variety of political issues and is widely regarded as being extremely cringe.[30] It is therefore perfect bot material.

The novel is a special edge case because of its deeply disturbing, offensive, long-form scenes, and in general for its atrocious writing. Here is an excerpt the bot frequently uses that is an example of the former. (Content warning: this is a fictionalized account of a police officer murdering a black child told sympathetically from the police officer’s point of view.)

Then he heard the voice.

“Hey, pig,” it said. The voice wasn’t deep. It was the voice of a child. And the kid stood outside the door of the quick mart, legs spread, arms hanging down by his sides. A cute black kid, wearing a Simpsons T-shirt and somebody’s old Converse sneakers and baggy jeans.

On his hip, stuck in those baggy jeans, was a pistol.

It looked like a pistol, anyway. But O’Sullivan couldn’t see clearly. The light wasn’t right. He could see the bulge, but not the object.

O’Sullivan put his flashlight back in his belt and put his hand back on his pistol, the greasy handle still warm to the touch.

“Stop right there, pig,” the kid said. His hand began to creep down toward his waistband.

O’Sullivan pulled the gun out of its holster, leveling it at the kid. “Put your hands above your head. Do it now!”

“Fuck you, honky,” the kid shot back. “Get the fuck out of my neighborhood.” Then he laughed, a cute kid’s laugh. O’Sullivan looked for sympathy behind those eyes, found none.

Oh, shit, O’Sullivan thought. Then he said, “Hands up. Right now.”

The kid laughed again, a musical tinkling noise. “You ain’t gonna shoot me, pig. What, you afraid of a kid?”

O’Sullivan could feel every breath as it entered his lungs. “No, kid, I don’t want to shoot you,” he said. “But I need you to cooperate. Put your hands above your head. Right now.”

The kid’s hand shifted to his waistband again. O’Sullivan’s hands began to shake.

“Get the fuck out of my neighborhood,” the kid repeated.

O’Sullivan looked around stealthily. Still nobody on the street. Totally empty. The sweat on his forehead felt cold in the night air. In the retraining sessions at the station, they’d told officers to remember the nasty racial legacy of the department, be aware of the community’s justified suspicion of police. Right now, all O’Sullivan was thinking about was getting this kid with the empty eyes to back the fuck off.

“Go on home,” he said.

“You go home, white boy,” said the kid. His hand moved lower.

Suddenly, O’Sullivan’s head filled with a sudden clarity, his brain with a preternatural energy. He recognized the feel of the adrenaline hitting. He wasn’t going to get shot on the corner of Iowa and Van Dyke outside a shitty convenience store in a shitty town by some eight-year-old, bleed out in the gutter of some city the world left behind. He had a life, too.

The gun felt alive in his hand. The gun was life.

The muzzle was aimed dead at the kid’s chest. No way to miss, with the kid this close, just ten feet away maybe. Still cloaked in the shadow of the gas station overhang.

“Kid, I’m not going to ask you again. I need you to put your hands on top of your head and get on your knees.”

“Fuck you, motherfucker.”

“I’m serious.”

The kid’s hand was nearly inside his waistband now.

“Don’t do that,” O’Sullivan said.

The kid smiled, almost gently.

“Don’t.”

The kid’s smile broadened, the hand moved down into the pants.

“Get the fuck out of my hood,” the kid cheerfully repeated. “I’ll cap your ass.”

“Kid, I’m warning you,” O’Sullivan yelled. “Put your hands above your head! Do it now...”

The roar shattered the night air, a sonic boom in the blackness. The shot blew the kid off his feet completely, knocked him onto his back.

O’Sullivan reached for his radio, mechanically reported it: “Shots fired, officer needs help at the gas station on Iowa and Van Dyke.”

“Ohgodohgodohgodohgod,” O’Sullivan repeated as he moved toward the body, the smoke rising from his Glock. He pointed it down at the kid again, but the boy wasn’t moving. The blood seeped through Homer Simpson’s face, pooled around the kid’s lifeless body. The grin had been replaced with a look of instantaneous shock. His hand had fallen out of his waistband with the force of the shooting.

In it was a toy gun, the tip orange plastic.

For a brief moment, O’Sullivan couldn’t breathe. When he looked up, he saw them coming. Dozens of them. The citizens of Detroit, coming out of the darkness, congregating. He could feel their eyes.

Officer Ricky O’Sullivan sat down on the curb and began to cry.⁶

⁶Because of Ben Shapiro’s history of weaponizing statistics, citations, and the credibility of other outlets, we have made the editorial decision not to cite directly to his content and we encourage others to do the same.

The novel is so absurd and offensive that it has become a meme in some parts of the internet. Several notable podcasts, including *Behind the Bastards*[31], have done readings.[30] This has led the Ben Shapiro bot itself to become popular in certain subreddits, such as /r/behindthebastards. Such users go on to purposefully summon the bot in other places.

As a nod to the novel’s notoriety in some communities, the bot is programmed to make inside jokes referencing *True Allegiance*. When a user tells the bot “good bot,” it responds with “Take a bullet for ya, babe,” a reference to the incredibly cringe terms of endearment used by the protagonist and his wife. This has led to users expressing confusion, which led to other users bringing them into the joke, welcoming them into the social group of people who find Ben Shapiro cringe.

What the bot links to

The bot’s messages contain a footer identifying itself as a bot. It also contains a link to a reddit wiki page explaining Ben Shapiro’s grift in more detail and a link where users can opt-out from future bot replies. While we are not aware of specific rules reddit enforces, a footer like this one is generally considered to be part of bot development best practices in the reddit community.

Reddit’s rules for bots

Reddit allows subreddits to ban bots from their communities. We note that the interventions we propose will not take place, and therefore can’t work, in any of the subreddits that have banned the bot. Of those subreddits, some are sympathetic to Ben Shapiro. We hope to intervene with users before they have become so involved in those communities that they don’t spend time on subreddits that have not banned the bot.

The Bot in Action/ Working

Here are some examples of the bot in action, to show the format of its comments and the content it includes and links to:

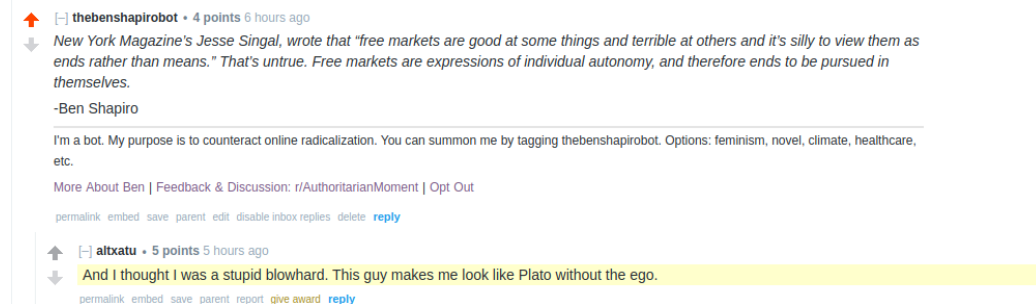


Figure 12: Funny comments mocking Ben Shapiro are an ideal outcome when the bot is invoked, as they expose other users to the disdain he's often met with, influencing their opinions of him.

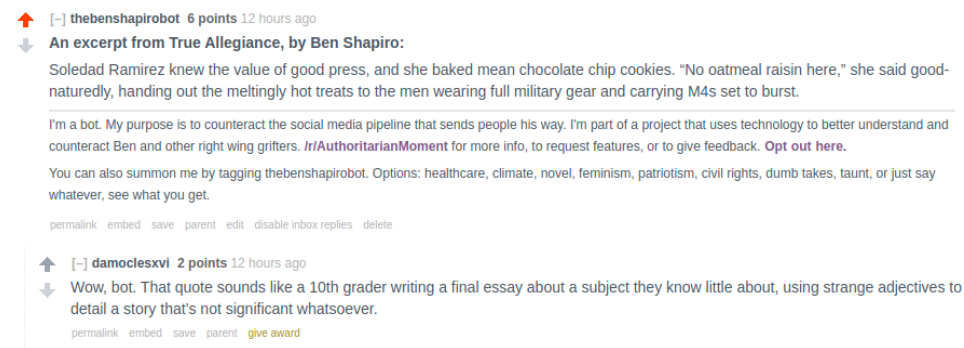


Figure 13: The content the bot provides is often surprising to users, hopefully causing them to re-evaluate any previously held opinions of Ben Shapiro.

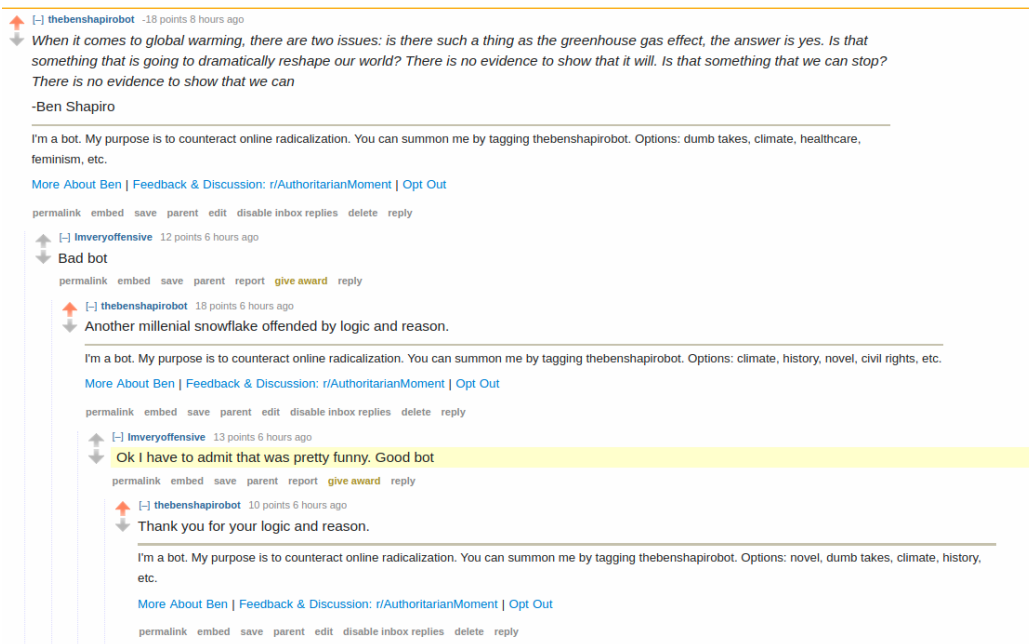


Figure 14: The humor the bot displays can sometimes win over initially hostile users.

Concluding remarks

Disinformation and hatred are significant societal problems. Ben Shapiro promulgates such content through a two-pronged approach: by laundering his image through the mainstream media and by targeting people likely to be vulnerable to his messaging in an attempt to grow his core fanbase. These people can then regurgitate his content more broadly into their social networks, which at some level intersect with the internet and the general population. The Ben Shapiro bot's interventions are like a firewall preventing the spread of Ben Shapiro's content through the broader system by preventing its initial penetration. The Ben Shapiro bot attempts to disrupt the mechanisms Ben Shapiro uses to disseminate his content by quoting Ben Shapiro directly in contexts likely to involve the same people being targeted by Ben Shapiro, exposing them to his less savory content and to the low opinion their

peers often have of Ben Shapiro. Quoting Ben Shapiro directly also counteracts the laundering of his image for people outside of his target audience, who may only be peripherally aware of Ben Shapiro. These interventions hopefully steer vulnerable users away from Ben Shapiro before they become full-throated fans, in particular by exposing them to their peers' judgment of him as "cringe".

Though the bot specifically targets Ben Shapiro, we hope to encourage alternative frameworks for combating disinformation and right-wing radicalization by examining and targeting dissemination systems that have been so successful, they have made the background noise of the internet Nazi memes.

Acknowledgements

Thank you to the Dangerous Speech Project, whose research into counter-speech helped inform early design decisions.

References

- [1] D. S. Project, "Dangerous speech: A practical guide," 2021. [Online]. Available: <https://dangerousspeech.org/guide/>
- [2] "Defense in depth (computing)," accessed on May 25, 2022. [Online]. Available: [https://en.wikipedia.org/wiki/Defense_in_depth_\(computing\)](https://en.wikipedia.org/wiki/Defense_in_depth_(computing))
- [3] M. Parks, "Outrage as a business model: How Ben Shapiro is using Facebook to build an empire." [Online]. Available: <https://www.npr.org/2021/07/19/1013793067/outrage-as-a-business-model-how-ben-shapiro-is-using-facebook-to-build-an-empire>

-
- [4] F. G. Kevin Rose, accessed on May 25, 2022. [Online]. Available: https://twitter.com/FacebooksTop10?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor
 - [5] Y. A. Foundation, “Ben Shapiro: Toughen up spoiled children,” 2015, accessed on April 24, 2022. [Online]. Available: <https://www.youtube.com/watch?v=MRqpXL8TxiUt=1444s>
 - [6] S. Tavernise, “Ben Shapiro, a provocative ‘gladiator,’ battles to win young conservatives,” *The New York Times*, Nov 2017, accessed on April 14, 2022. [Online]. Available: <https://www.nytimes.com/2017/11/23/us/ben-shapiro-conservative.html>
 - [7] “US commentator clashes with BBC’s Andrew Neil.” [Online]. Available: <https://www.youtube.com/watch?v=6VixqvOcK8E>
 - [8] “Ben shapiro on the ‘evolution of american values,’” this contains an archive link to avoid citing directly to Ben Shapiro’s work. Because of Ben Shapiro’s history of weaponizing statistics, citations, and the credibility of other outlets, we have made the editorial decision not to cite directly to his content and we encourage others to do the same. [Online]. Available: <https://archive.ph/Vhice>
 - [9] “Pbs frontline: Ben shapiro interview,” this contains an archive link to avoid citing directly to Ben Shapiro’s work. Because of Ben Shapiro’s history of weaponizing statistics, citations, and the credibility of other outlets, we have made the editorial decision not to cite directly to his content and we encourage others to do the same. [Online]. Available: <https://archive.ph/cR3Ao>
 - [10] “Politico playbook: The real reason most republicans opposed impeachment,” this contains an archive link to avoid citing directly to Ben Shapiro’s work. Because of Ben Shapiro’s history of weaponizing statistics, citations, and the credibility of other outlets, we have made the editorial decision not to cite directly to his content and we encourage others to do the same. [Online]. Available: <https://archive.ph/17NY2>
 - [11] “Stacey abrams wants a divided america,” this contains an archive link to avoid citing directly to Ben Shapiro’s work. Because of Ben Shapiro’s

history of weaponizing statistics, citations, and the credibility of other outlets, we have made the editorial decision not to cite directly to his content and we encourage others to do the same. [Online]. Available: <https://archive.ph/W03Gy>

- [12] M. F. Suarez, “Cafecito talk with Ben Shapiro.” [Online]. Available: https://www.youtube.com/watch?v=n5usOlQnR_wab_channel=MayorFrancisSuarez
- [13] [Online]. Available: <https://www.post-gazette.com/opinion/Op-Ed/2022/05/23/media-lie-conservatives-replacement-theory-buffalo-shooter-ben-shapiro/stories/202205230012>
- [14] Archive link: <https://archive.ph/I7rgC>.
- [15] Archive link: <https://archive.ph/rbVNg>.
- [16] S. F. Kiesling, “Dude,” *American Speech*, vol. 79, no. 3, pp. 281–305, 08 2004. [Online]. Available: <https://doi.org/10.1215/00031283-79-3-281>
- [17] M. Bucholtz, “The whiteness of nerds: Superstandard english and racial markedness,” *Journal of Linguistic Anthropology*, vol. 11, pp. 84–100, 2001.
- [18] Google, “Google books ngram viewer,” 2022. [Online]. Available: https://books.google.com/ngrams/graph?content=facts+don%27t+care&year_start=1800year_end=2019corpus=26smoothing=3direct_url=t1%3B%2Cfacts%20do%20not%20care%3B%2Cc0t1%3B%2Cfacts%20do%20not%20care%3B%2Cc1
- [19] Psychological_Web837, 2022, accessed on April 22, 2022. [Online]. Available: https://www.reddit.com/r/benshapiro/comments/u96k3h/comment/i5pt65i/?utm_source=shareutm_medium=web2xcontext=3
- [20] Benny-Boi135, 2022, accessed on April 22, 2022. [Online]. Available: https://www.reddit.com/r/benshapiro/comments/u96k3h/comment/i5rj1o3/?utm_source=shareutm_medium=web2xcontext=3
- [21] Fuzzy_Education4705, 2022, accessed on April 23, 2022. [Online]. Available: https://www.reddit.com/r/benshapiro/comments/s1im6c/finally_it_happened/

- [22] MagicGeoff, 2022, accessed on April 23, 2022. [Online]. Available: https://www.reddit.com/r/benshapiro/comments/slim6c/comment/hvu3f77/?utm_source=shareutm_medium=web2xcontext=3
- [23] nebulouslurker, 2022, accessed on April 23, 2022. [Online]. Available: https://www.reddit.com/r/benshapiro/comments/slim6c/comment/hvv2xmm/?utm_source=shareutm_medium=web2xcontext=3
- [24] CollectionPotential, 2022, accessed on April 23, 2022. [Online]. Available: https://www.reddit.com/r/benshapiro/comments/slim6c/comment/hsa3amu/?utm_source=shareutm_medium=web2xcontext=3
- [25] T. M. R. w/ Sam Seder, 2021, accessed on April 24, 2022. [Online]. Available: https://www.youtube.com/watch?v=yNfDBWYgrH0ab_channel=TheMajorityReportw%2FSamSeder
- [26] oliSUNvia, 2021, accessed on April 24, 2022. [Online]. Available: https://www.youtube.com/watch?v=R3r01BruBokab_channel=oliSUNvia
- [27] J. Theodora, 2021, accessed on April 24, 2022. [Online]. Available: https://www.youtube.com/watch?v=w4vbtBmQ4IUab_channel=JasmyneTheodora
- [28] R. J. Gallagher, M. R. Frank, L. Mitchell, A. J. Schwartz, A. J. Reagan, C. M. Danforth, and P. S. Dodds, “Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts,” *EPJ Data Science*, vol. 10, no. 1, jan 2021. [Online]. Available: <https://doi.org/10.1140%2Fepjds%2Fs13688-021-00260-3>
- [29] A. Reagan, “labmtsimple documentation,” 2018, accessed on April 25, 2022. [Online]. Available: <https://buildmedia.readthedocs.org/media/pdf/labmt-simple/latest/labmt-simple.pdf>
- [30] R. Evans, K. Stoll, and C. Johnston, “Ben shapiro’s terrible book (as read on behind the bastards),” this is a list of the relevant episodes curated by user geekcreate. [Online]. Available: <https://www.podchaser.com/lists/ben-shapiros-terrible-book-as-read-on-behind-the-bastards-107a9FVacT>

- [31] R. Evans, “Behind the bastards.” [Online]. Available:
<https://www.iheart.com/podcast/105-behind-the-bastards-29236323/>