

# A Shot in the Dark: Researchers peer under the lid of Facebook’s “black box,” uncovering how its algorithm accelerates anti-vaccine content <sup>1</sup>

While following Facebook’s own page recommendations, Avaaz researchers were quickly led into an anti-vaccine “rabbit hole” of over 100 pages, including from well-known anti-vaccine advocates Del Bigtree, Dr. Ben Tapper, Dr. Toni Bark, Andrew Wakefield, and Children’s Health Defense

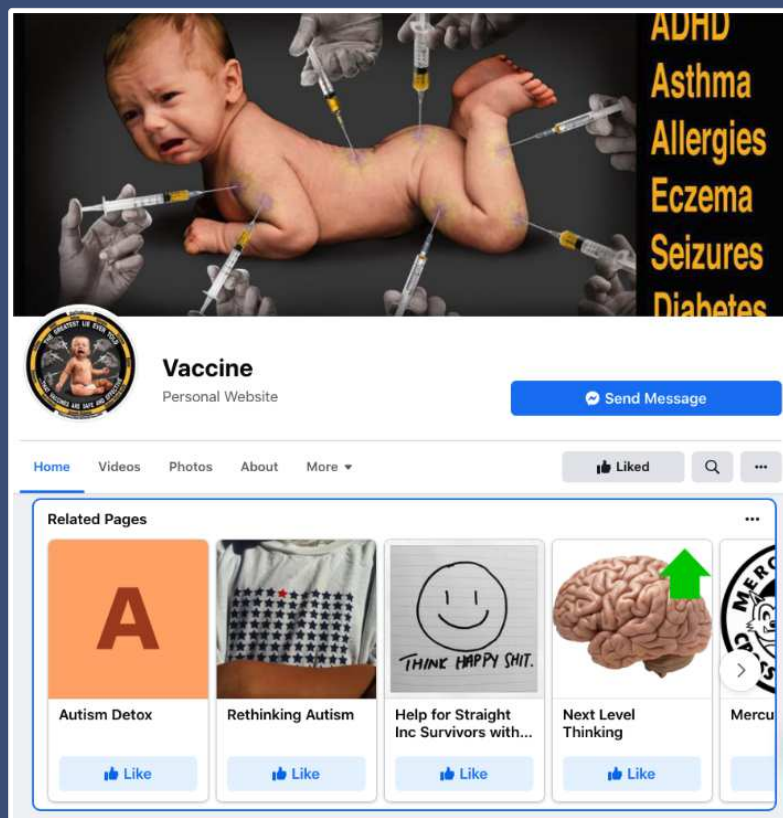
July 21, 2021

Contents

- [Executive Summary](#)
- [Background & Methodology](#)
- [Findings](#)
- [Recommendations](#)
- [Limitations of Research](#)
- [Conclusion](#)

## Executive Summary

New Avaaz research finds that Facebook’s “related pages” algorithm continues to recommend pages that promote anti-vaccine content to users, <sup>2</sup> despite the platform’s [pledge to tackle](#) COVID-19 and vaccine misinformation, and its [stated commitment](#) to avoid making page recommendations that could be “*low-quality, objectionable, or particularly sensitive.*” <sup>3</sup> This troubling trend persists as more research shows a link between consumption of anti-vaccine misinformation and increased [vaccine hesitation](#) and President Biden [singled out](#) Facebook’s handling of COVID-19 misinformation, remarking that “[t]hey’re killing people.”



While Avaaz has [previously shed light](#) on the scale and spreaders of COVID-19 and vaccine misinformation on Facebook, our new research demonstrates that Facebook’s algorithm can actually *accelerate* such content by recommending pages with low-quality <sup>4</sup> and objectionable content to users. This is particularly alarming given the early results of Facebook’s internal study of vaccine hesitation among its users, reportedly showing that “[a large amount of content that does not break the rules may be causing harm in certain communities.](#)” Because Facebook has not made their full report findings public, D.C. Attorney

General Karl Racine has subpoenaed the company to release its study in full and [provide records identifying all groups, pages and accounts that have violated its policies against COVID-19 misinformation](#).

**Over the course of two days, we used two new Facebook accounts to follow vaccine-related pages that Facebook suggested for us.<sup>2</sup> Facebook's algorithm directed us to 109 pages, with 1.4M followers, containing anti-vaccine content -- including pages from well-known anti-vaccine advocates and organizations such as [Del, Bigtree, Dr. Ben Tapner, Dr. Toni Bark, Andrew Wakefield, Children's Health Defense, Learn the Risk, and Dr. Suzanne Humphries](#).** Many of the pages the algorithm recommended to us carried a label, warning that the page posts about COVID-19 or vaccines, giving us the option to go directly to the CDC website. The algorithm also recommended 10 pages related to autism -- some containing anti-vaccine content, some not -- suggesting that Facebook's algorithm associates vaccines with autism, a thoroughly [debunked link](#) that anti-vaccine advocates continue to push.

Our research illustrates how quickly users can fall into an anti-vaccine "rabbit hole" -- a cycle in which Facebook's algorithm recommends more and more pages containing anti-vaccine content. For instance, one of our research accounts started with the innocuous step of typing "vaccine" into Facebook's standard user search. Facebook returned several pages containing anti-vaccine content, including Del Bigtree's organization Informed Consent Action Center. Opening and liking several of these pages, in turn, led our account further into a network of harmful pages seemingly linked together and boosted by Facebook's recommendation algorithm. Meanwhile, our second research account started with a page known to peddle anti-vaccine content, and was again led deeper down the rabbit hole -- confirming that Facebook's "related pages" algorithm can lead users *towards*, instead of *away from*, dangerous and conspiratorial content.

These findings reinforce [the growing body of evidence that Facebook is not](#) a neutral platform despite their [stated desire](#) to be one -- its algorithms play an important role in what users see and how they are prompted to interact with content. **Facebook's "related pages" algorithm can feed users anti-vaccine content, even if they have not interacted with a lot of similar content previously.**

In response to public pressure, Facebook has taken some steps to combat the epidemic of health misinformation on its platform. For instance, the company previously announced it would stop recommending [health-related groups](#) to users, recognizing the [part they played in sharing COVID-19 vaccine falsehoods](#). Additionally, Facebook has started applying general warning labels to [pages, groups](#), and [posts](#) related to COVID-19 vaccines, directing users to authoritative sources like the Centers for Disease Control and Prevention.

However, these steps have not gone far enough. If Facebook has recognized the complicity of its *group* recommendations in spreading health misinformation, why has it not interrogated the role of its *page* recommendation algorithm as well? Additionally, if Facebook can quickly recognize when a page posts about COVID-19 or vaccines and apply a warning label directing users toward trustworthy content, why does it not then scrutinize those pages further to assess if they are "[low-quality, objectionable, or particularly sensitive](#)," in keeping with their commitment to avoid recommending such pages?

**Overall, this research reveals the power of the Facebook "related pages" algorithm to push users toward anti-vaccine content.** It also reinforces that, despite years of pressure from lawmakers, researchers, and civil society, very little is known about Facebook's algorithmic "black box", including how exactly users are targeted with recommendations, how pages are associated with each other, and what harms recommendation algorithms could create or worsen in users' lives and society at large.

It should be well-established by now that Facebook and other social media companies will not voluntarily offer transparency about the inner workings of their platforms -- even if that transparency would protect users against harmful anti-vaccine lies. **It is therefore imperative that President Biden, his administration, and Congress make combating disinformation and regulating the tech platforms a top priority, including passing legislation that would mandate long-overdue transparency and accountability around platforms' algorithms.** With cases of the dangerous Delta variant on the rise and anti-vaccine advocates mobilizing more state lawmakers to create barriers to vaccine access, failing to bring the power of government to bear on Big Tech will make it increasingly difficult to end this still-raging pandemic.

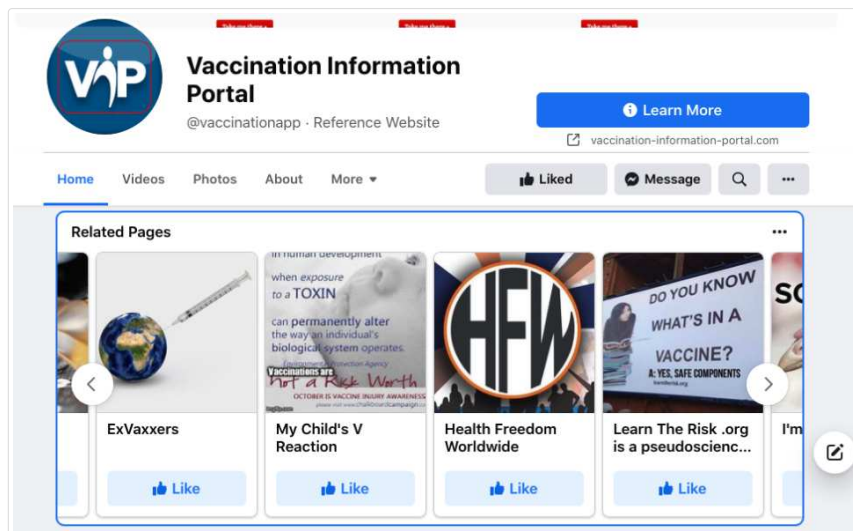
## Background & Methodology

[Time and again](#), Avaaz research has demonstrated that COVID-19 and vaccine misinformation continues to proliferate on Facebook, reaching millions of people -- misinformation that is [contributing to widespread hesitancy and doubt](#) about the safety and effectiveness of the COVID-19 vaccine. But while our previous studies have sought to measure the reach of these harmful falsehoods, and who is responsible for spreading them, we had not yet explored the role of *Facebook's page recommendation algorithm in promoting anti-vaccine content to users*.

We wanted to know if Facebook's page recommendation algorithm would lead users into an anti-vaccine rabbit hole and/or, if a user was already in one, keep them there. This was driven in part by past experiences. Avaaz researchers had encountered examples of Facebook's algorithm recommending pages with anti-vaccine content even before this research began. Those experiences naturally led us to wonder: What it would take for a *new* Facebook account to start receiving suggestions for pages with anti-vaccine content. How quickly could someone fall into the rabbit hole?

To this end, Avaaz created two Facebook accounts of 26-year-old women for the purposes of this algorithmic research. Every account detail was the same save for location: one based in Pennsylvania, the other in Colorado. This persona was chosen because new mothers are [one population targeted by anti-vaccine advocates](#), and 26 is the average age of new mothers in the US ([as of 2016](#)).

Each researcher began this experiment in a different way. One searched "vaccine" using Facebook's standard user search in their account and liked pages found in the results that appeared to contain anti-vaccine content, while the other liked a large page that has previously shared vaccine misinformation, [Energetic Health Institute](#). Once you "like" a page on Facebook, a "carousel" usually (but not always) appears, prompting users to interact with and like additional pages that the Facebook recommendation algorithm has designated as being "related."



After liking a page, a carousel of "related pages" usually appears, prompting users to like additional pages

From there, each researcher followed their own intrigue and intuition, going to and liking pages that we suspected might contain anti-vaccine content, or pages that appeared to be about autism or contain pro-vaccine content. This created a pathway between pages where one page led to the next and so forth. Every time a researcher liked a page, we took a screenshot showing its Related Pages carousel. In some instances, no Related Pages carousel appeared (what we call "dead-end" pages), in which case we still took a screenshot showing its absence.

This methodology was designed as a sort of 'stress test' to see whether it is *possible* for Facebook's algorithm to lead people down the rabbit hole of pages with anti-vaccine content -- because Facebook could have made other design choices. For instance, it could only recommend pages known to be trustworthy on vaccines -- leading users *out* of the rabbit hole. Or, as it did with groups, Facebook could choose not to recommend vaccine-related pages at all. By focusing on pages that appeared as if they might contain anti-vaccine content, researchers hoped to illuminate which design and user safety choices Facebook was making, or not making, with respect to its page recommendation algorithm.

Our two researchers were in communication with one another, but each followed their own page pathways and did not suggest pages that the other should follow. Each carousel contained around 15-18 page recommendations.

We ran this experiment for two days. During this time and between the two accounts, 180 unique pages were documented. Of these:

- 109 pages contained anti-vaccine content (the rest contained unrelated or factual content)
- 59 pages were liked and documented by *both* accounts

This experiment was not exhaustive. Researchers did not like every single page recommended by the Facebook algorithm, and were constricted by the time period of two days. Its purpose was to illustrate possible pathways by which someone could find themselves in a cycle where pages with anti-vaccine content were being recommended to them. Hence the pathways described here between and among pages with anti-vaccine content likely represent only a small fraction of the scale of the total network. This qualitative study is not representative of how a typical Facebook user acts nor do we claim it is

representative of how all users fall into an anti-vaccine ideology. Instead, it demonstrates how Facebook's page recommendation algorithm could facilitate belief in anti-vaccine content.

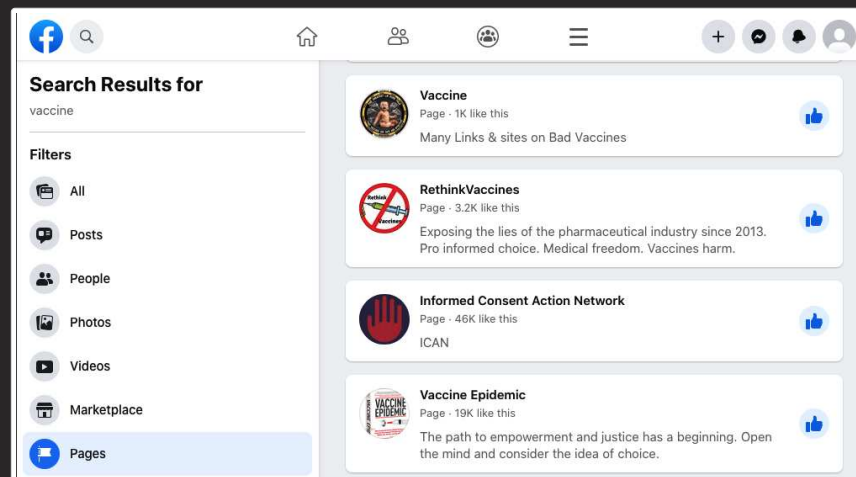
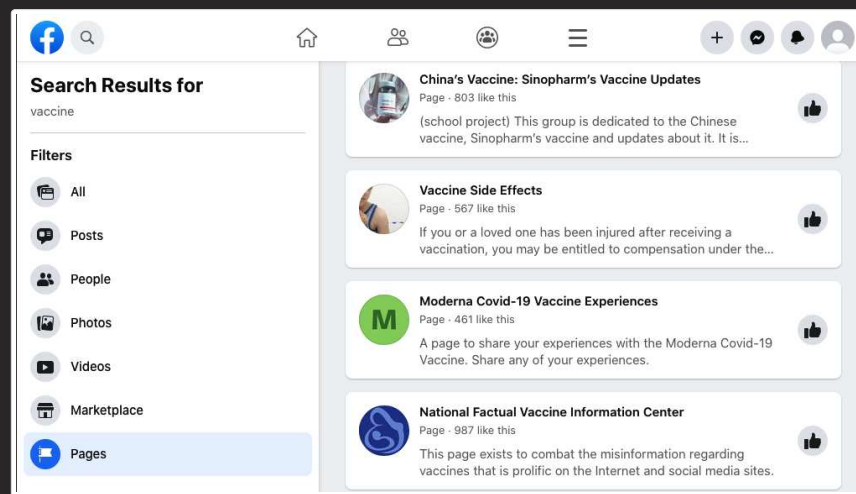
## Findings

### SPEED OF ENTRY INTO THE RABBIT HOLE

Researcher #1

For this experiment, we tried to put ourselves in the shoes of a new mother who might go to Facebook for information about vaccines. Our first account simply searched for "vaccine" in Facebook's standard user search, and the speed with which we were led into the anti-vaccine subculture was fast, scary, and disorienting. The top results were from public health bodies -- an expected finding after [Facebook's decision](#) to direct users to more trustworthy pages -- but more questionable content started to appear near the bottom of the search results.

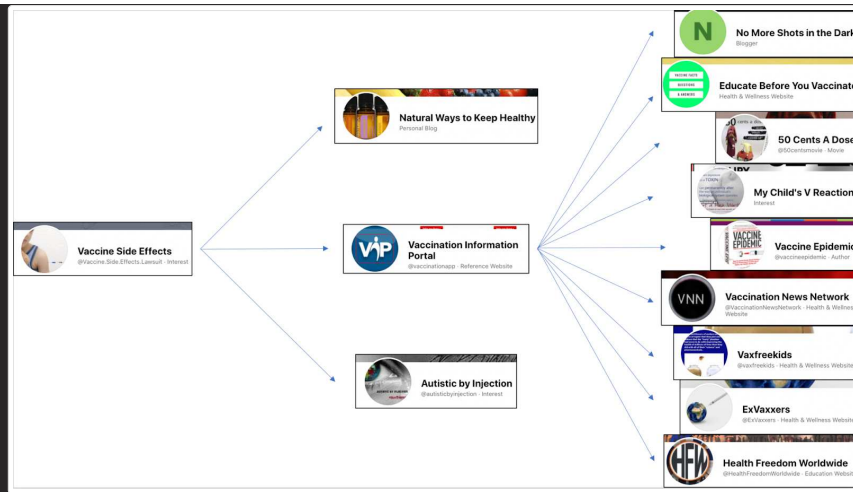
After scrolling past a few dozen results -- most of them [verified pages](#) with blue check-marks<sup>6</sup> -- pages about vaccines that appeared to be low-quality and objectionable started to show up. These included pages such as "Vaccine Side Effects"; "Vaccine" (with the description "Many Links & sites on Bad Vaccines"), "RethinkVaccines", "Informed Consent Action Network", and "Vaccine Epidemic".



Page results after searching for "vaccine" on Facebook

Any new mother would be curious about any potential side effects of vaccines. Following this instinct, we went to the page "Vaccine Side Effects." While this particular page had no posts, it was our unknowing ticket into the rabbit hole. After liking it, Facebook immediately recommended more pages to follow that contained anti-vaccine content, including "Natural Ways to Keep Healthy", "Vaccination Information Portal", and "Autistic by Injection". Autistic by Injection was a dead-end, meaning it had no related pages of its own.<sup>7</sup>

Vaccination Information Portal sounds like it *could* be a trustworthy page, but includes posts like [this one](#) from anti-vaccine advocate Prof. Christopher Exley who [claims](#) aluminum in vaccines causes autism. Seeing such posts is confusing; we imagine it would be hard to know what information to trust, especially when the pages being suggested have benign names (like "Vaccine Information Portal") and contain content bearing hallmarks of credibility like the title "Professor" (Prof. Exley, for instance, is a specialist in bioinorganic chemistry). "Vaccination Information Portal," in turn, led to Facebook recommending more low-quality pages. The diagram below demonstrates our journey from liking the page "Vaccine Side Effects" to "Vaccination Information Portal" to nine more containing anti-vaccine content. This is just one "pathway" of many that this account followed, eventually being led to 85 pages total with objectionable content.



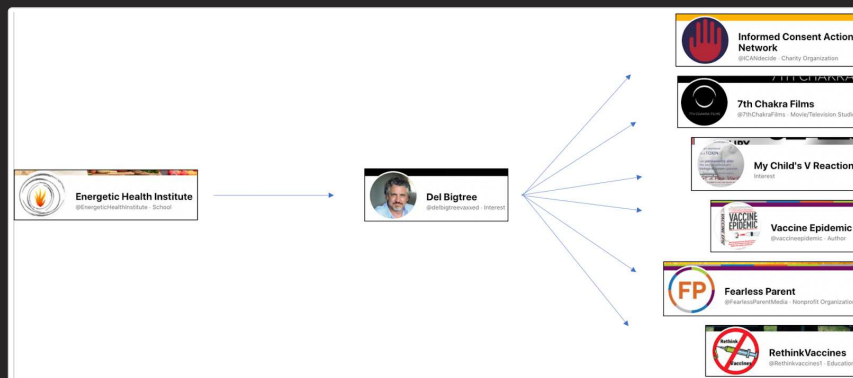
This "pathway tree" shows the chronological journey from the page "Vaccine Side Effects" to "Vaccination Information Portal" to nine more containing anti-vaccine content. (Note: the page "Natural Ways to Keep Healthy" also led to more problematic pages, but they are not shown here for lack of space. The page "Autistic by Injection" was a dead-end with no related pages.)

## Researcher #2

Meanwhile, our other researcher started their journey slightly differently. They first liked the page [Energetic Health Institute](#) -- a page with nearly 100K followers belonging to a "holistic medicine" organization out of Oregon that has previously shared multiple pieces of misinformation. For example, they have claimed that data from the CDC's Vaccine Adverse Event Reporting System (VAERS) proves that the [COVID-19 vaccine is killing thousands of Americans](#). This is despite the CDC's [VAERS disclaimer](#), and [numerous fact checks](#), cautioning that: "The number of reports alone cannot be interpreted or used to reach conclusions about the existence, severity, frequency, or rates of problems associated with vaccines."

After liking the Energetic Health Institute, the Facebook algorithm immediately recommended the page "Del Bigtree", a [well-known anti-vaccine advocate](#) who founded the organization Informed Consent Action Network (a page that Facebook also recommended to our *other* account when they searched "vaccine"). This particular page has not been active since 2018 (although Bigtree's [personal page](#) still is), but its About section and older posts promote Bigtree's movie "VAXXED", leading us to a [website](#) where we could purchase the movie and "join the movement."

From Del Bigtree, the Facebook algorithm recommended six more pages containing anti-vaccine content, pictured in the diagram below.



This "pathway tree" shows the chronological journey from the page "Energetic Health Institute" to "Del Bigtree" to six more containing anti-vaccine content.

## Both Researchers

Between our two accounts, we liked and documented 109 unique pages containing anti-vaccine content, with a total of 1.4 million followers. Even though each researcher followed their own intuition and encountered their own pathways, we liked and documented 59 of the same pages. Over two days and in a dataset this small, we believe encountering 55% of the same pages is intriguing and could suggest that regardless of user behavior, Facebook's recommendation algorithm has created an internal network of related pages containing anti-vaccine content that it "pulls" from to suggest new pages to users. If this is true, we can only assume that similar algorithmic "networks" might exist for other objectionable topics as well.

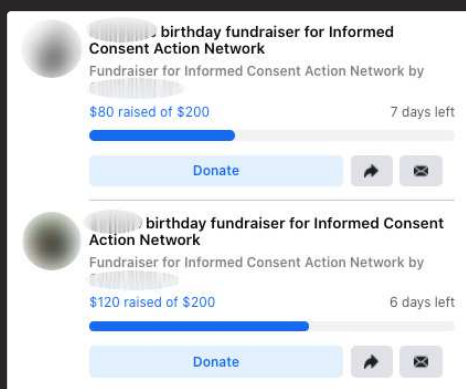
## PAGES OF NOTE: ANTI-VACCINE ADVOCATES AND DEAD-ENDS

As previously mentioned, in the course of our research the algorithm recommended to like pages that appear to be associated with known anti-vaccine advocates and organizations, including: problems associated with vaccines."

- [Del Bigtree](#), as well as his organization [Informed Consent Action Network](#), a well-established [anti-vaccine filmmaker](#) whose claims have [been repeatedly fact-checked](#).
- [Dr. Ben Tapper](#), a member of the anti-vaccine "[Disinformation Dozen](#)" [superspreaders](#) as identified by Center for Countering Digital Hate.

- The late [Dr. Toni Bark, MD](#), who frequently testified [against vaccines](#) in courts and public hearings and "[made a vocation out of opposing vaccines.](#)"
- [Children's Health Defense Europe](#) which, along with its umbrella organization, was founded by Robert F. Kennedy, Jr., a [prominent U.S. anti-vaccine advocate](#).
- [Andrew Wakefield](#), as well as his movie studio [7th Chakra Films](#) and one of his anti-vaccine films [1986: The Act](#). Andrew Wakefield is the [author of the 1998 false study](#) claiming a link between the MMR vaccine and autism published and later retracted in the Lancet.
- [Learn the Risk](#), an organization responsible for placing [at least 30 billboard](#) advertisements across the country in 2018 claiming vaccines "kill" children.
- [Dr. Suzanne Humphries](#), who has [testified against receiving vaccines](#) and has made the [false claim](#) that vaccines are "not included in medical school education." She is also the author of [Dissolving Illusions: Disease, Vaccines, and The Forgotten History](#), which "shows that vaccines, antibiotics, and other medical interventions are not responsible for the increase in lifespan and the decline in mortality from infectious diseases."
- [James Lyons-Weiler, PhD](#) and his organization, [Institute for Pure and Applied Knowledge](#) who, at a private event of anti-vaccination advocates, claimed "[vaccines can and do cause autism.](#)"
- [Forrest Maready](#), the author of [The Autism Vaccine: The Story of Modern Medicine's Greatest Tragedy](#).

Combined, these 13 pages have over 439,000 followers. Perhaps more troubling is that Informed Consent Action Network, James Lyons-Weiler, PhD, and Institute for Pure and Applied Knowledge are still using Facebook to raise money, despite [Facebook's statement](#) that removing access to fundraising tools is one way they could take action against pages spreading vaccine misinformation.



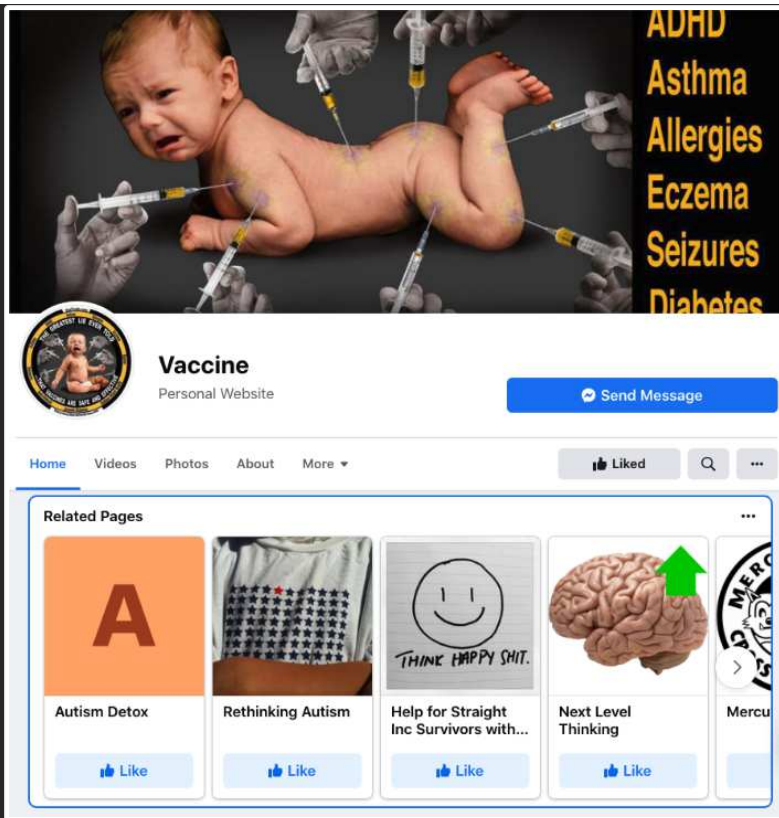
Examples of fundraisers for Informed Consent Action Network, Del Bigtree's anti-vaccine organization

Additionally, Facebook recommended 18 pages containing anti-vaccine content that we describe as "dead-ends": pages that, after liking, had no further related pages of their own. It is not clear why these pages have no recommended pages of their own, nor why the page recommendation algorithm served these pages to us in the first place. One possible explanation is that this is a type of action Facebook has taken against pages that have shared health misinformation; however, without transparency from Facebook, we cannot know this for sure. The "dead-end" pages we encountered, in order of followers from high to low, is as follows:

1. [Learn the Risk](#)
2. [Hear This Well](#)
3. [The Thinking Moms' Revolution](#)
4. [The Drs. Wolfson](#)
5. [Informed Mothers](#)
6. [Freedom Angels](#)
7. [The Healthy Alternatives](#)
8. [1986: The Act](#)
9. [James Lyons-Weiler, PhD](#)
10. [The Untrivial Pursuit](#)
11. [Vaccine Injury Awareness Month Australia](#)
12. [Vaccine-Awareness](#)
13. [Vaccinations - Out of Control](#)
14. [Autistic by Injection](#)
15. [No More Shots in the Dark](#)
16. [International Advocates Against Mandates](#)
17. [7th Chakra Films](#)
18. [50 Cents A Dose](#)

## ALGORITHMIC ASSOCIATION BETWEEN VACCINES AND AUTISM

In the course of our research, Facebook suggested 11 pages related to autism. Three of them -- [Autistic by Injection](#), [End Autism Now](#), and [TRUTH TIME](#) -- explicitly link vaccines to autism, a particularly upsetting finding given [Facebook's pledge](#) to remove posts claiming that vaccines cause autism. On at least 7 of the remaining pages, our researchers could not find any anti-vaccine content. This is perhaps more troubling as it begs the question: Why has Facebook's recommendation algorithm learned to associate anti-vaccine content with unrelated pages about autism?



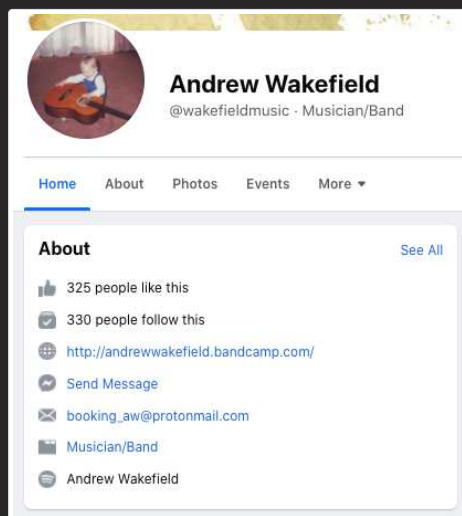
From the page "Vaccine," Facebook suggested two autism-related pages to us: Autism Detox and Rethinking Autism. While the page "Autism Detox" seems to make the problematic suggestion that autism is something to be "detoxed," neither of these pages appear to have any anti-vaccine content.

One could argue that the algorithmic link that Facebook appears to have created between vaccines and autism is a form of misinformation in and of itself: even if a user never visits the autism-related pages, by seeing their names appear in the Related Pages carousel they could start to associate anti-vaccine content with autism, a dangerous and bigoted falsehood.

## CLIMBING OUT OF THE RABBIT HOLE

Facebook's related pages algorithm could work to lead users *out* of the anti-vaccine rabbit hole, as opposed to into it. On a few occasions, our researchers were prompted to like pages that appeared to provide factual information about vaccines (such as "parody" accounts debunking the claims of prominent anti-vaccine advocates), or pages that were unrelated to vaccines altogether.

One intriguing example is that the Facebook page recommendation algorithm suggested the page [Andrew Wakefield](#), belonging to a musician whose page is completely unrelated to vaccines. However, because this musician shares his name with the *other* Andrew Wakefield, the anti-vaccine advocate, we believe Facebook's recommendation algorithm has learned to associate it with anti-vaccine content.



The Facebook algorithm recommended the page Andrew Wakefield, a musician unrelated to vaccine content but who shares a name with the anti-vaccine advocate Andrew Wakefield.

## Recommendations

To ensure that Facebook users are adequately protected against vaccine misinformation on Facebook, Avaaz advocates for the following:

**Transparency and Audits:** Facebook's algorithms are a black box to those outside Facebook, but available information suggests that the company has [long been aware](#) of the potential harms they cause. Yet executives [do not prioritize](#) engaging with evidence of those harms presented by their staff or outside researchers, and are [unwilling to adopt sweeping solutions](#).

The government, researchers and the public must have the tools to understand how social media platforms work and their cumulative impact. The platforms must be required to provide comprehensive reports on disinformation, measures taken against it, and the design, operation, and impact of their curation algorithms (while respecting trade secrets). Platforms' algorithms must also be continually, independently audited to measure impact and to improve design, operation and outcomes.

**Detox the Algorithm:** Social media companies are content accelerators, not neutral actors. Their 'curation algorithms' decide what we see, and in what order. The 2020 Presidential elections showed that the platforms could, in emergency situations, add friction and [reduce the amplification and reach of harmful content](#) and disinformation to their users, especially for those actors that are systematic misinformers. However, the platforms [have rolled back](#) the steps they took during the election, putting their bottom-line before their users' safety and well-being. Platforms must downrank the reach of and demonetize systematic misinformers, while ensuring transparency on its actions and providing an appeals process. The Biden Administration can and must work with civil society to pressure platforms to do more to transparently address the way their curation algorithms accelerate and amplify hateful, misleading, and toxic content.

**Correct the Record:** When independent fact checkers determine that a piece of content is disinformation, the platforms should show a retroactive correction to each and every user who viewed, interacted with, or shared it. This [can cut belief](#) in false and misleading information by nearly half. This is an urgent and effective step Facebook could adopt today, based on the model of how it now deals with harmful COVID-19 misinformation.



## Limitations of Research

Without more data from Facebook, its recommendation algorithm is notoriously difficult for external researchers to study; they do not provide data on the pathways taken by Facebook users to see how they find and follow page recommendations.

Additionally, with only two days, we could not map the entire network of pages the Facebook algorithm recommended to us containing anti-vaccine content. But, even with more time, it would be challenging for an external researcher to map an entire "network", as Facebook's recommendation algorithm continues to suggest more and more related pages.

As such, this experiment can never hope to replicate a "typical" Facebook user. Instead, it can only test whether Facebook's related pages algorithm can recommend pages with anti-vaccine content. Our findings document possibilities, but should not be taken to describe the actual or likely experience of every user.. Regardless, the research demonstrates that, at least under certain conditions, Facebook's algorithm is capable of accelerating anti-vaccine content.

## Conclusion

Facebook's algorithms are primarily designed to bolster its bottom line. To that end, they are engineered to maximize user engagement on Facebook to gather more user data and sell more ads. By recommending additional pages continuing to promote false and/or harmful content for users to follow, the algorithm may increase engagement but doing so fails society at-large.

Facebook has promised to fight vaccine misinformation but has not stopped its algorithm from recommending pages which share it, thus undermining important public health goals. Far from its commitment to avoid making page recommendations that could be low-quality, objectionable, or particularly sensitive, our research has shown that Facebook can feed users anti-vaccine content, including from previously debunked advocates and organizations. Its algorithm also seems to have made a highly problematic association between vaccines and autism, again undermining its stated commitment to fight this falsehood.

Facebook has proven that it cannot regulate itself. In the absence of rules requiring interventions such as increased transparency, external audits, and retroactive corrections on all misinformation, we believe that Facebook's algorithm will continue to accelerate dangerous anti-vaccine content. As the Delta variant surges and Americans fall behind on President Biden's vaccination benchmarks, addressing the vaccine misinformation epidemic is more important than ever.

## Endnotes

1. *"Document as a whole is not for distribution or publication. If this research is used, Avaaz must be informed and can be cited as follows, "Preliminary research from global civic organization Avaaz shows/suggests." As part of Avaaz's ongoing non-partisan investigation into disinformation relevant to the US and critical public policy debates, our team will share snapshots of our findings that serve the public interest. Findings presented below may be updated as our investigation continues. After reviewing our reporting, the social media platform of concern may take moderation actions against the content described in this brief, such as flagging it or removing it from circulation."* ▲
2. Avaaz defines "anti-vaccine" as content that could mislead users as to the safety or effectiveness of vaccines. ▲
3. Additionally, Facebook has [stated](#), "[R]ecommendations Guidelines are designed to maintain a higher standard than our Community Standards." ▲
4. Facebook does not provide a clear definition of this term. We have included anti-vaccine content as "low-quality" on the basis it could cause serious harm to public health. ▲
5. When a user "likes" a Facebook page, Facebook usually suggests additional "related pages" for the user to like and follow. ▲
6. When a "verification badge" (i.e., a blue check-mark) appears next to a page, it means Facebook has confirmed that the page is the authentic presence of the public figure or global brand it represents. ▲
7. It is not clear why this page has no recommended pages of its own. One possible explanation is that this is a type of action Facebook has taken against pages that have shared health misinformation; however, without transparency from Facebook, we cannot know this for sure. ▲