

The Mechanics and Psychology Behind the Social Dilemma

Jeff Seibert

21-27 minutes

The Social Dilemma, a years-long passion project from [Jeff Orlowski](#) and the Exposure Labs team, just premiered on Netflix this past week, and to my amazement and delight it has quickly jumped onto the Top 10 list this weekend.

The film masterfully intermixes drama with sit-down interviews with myself and a dozen other former social media execs, in hopes of communicating the (often subconscious) consequences of the world's growing dependence on social media for news and entertainment.

It's fantastic that this message is beginning to resonate broadly.

For those within the tech industry, of course, questions abound. In just the past few days, I've been contacted by product managers, engineers, designers, and other tech execs in Silicon Valley and beyond, curious to better understand what exactly is at play here, and what if anything they can do about it.

There is also some confusion about how the film came about, who was involved, and how it relates to other hot topics in tech over the past decade.

To make the film as widely-accessible as possible, the Exposure Labs team obviously kept everything relatively high-level. If you're interested in diving beneath the surface, and understanding the product mechanics and the human psychology at play, here is my own take on what is happening under the covers.

First, some key points to level-set on:

1. This post is blame-free and not targeted at anyone. These challenges are structural, not the result of any one person, team, or company. These companies are full of amazing, smart, passionate, ethical folks, and everyone I ever interacted with (I was there!) wanted deeply to do what was best for people around the world.
2. Where we are today is not due to anyone's lack of execution. In fact, it was excellent execution that brought us here: the algorithms that power today's platforms succeeded beyond imagination, and engineering-product-design teams iterated rapidly to successfully achieve company goals. People were promoted and widely recognized because of this work, it was not hidden in the shadows, and there is nothing confidential about these topics.
3. Today, these issues are a lot less controversial than they were when I was interviewed for the film in early 2018, and I heard from the production team that it was very challenging to convince anyone at that time to go on the record about these topics. I get it—I was personally hesitant to be interviewed, though I'm now glad Jeff Orlowski persuaded me to do it.
4. I am sure there were many other people thinking about the same things and who would have made great additions to the film. I know the production team reached

out to a LOT of people to try to convince them to join, with mixed success. I'm sure there are additional folks that they missed.

5. There is a lot of technical subtlety to these challenges that is hard to express in writing, much less in film. I also don't pretend to know all of it, so I hope this post is the *beginning* of a productive conversation within the tech industry on how to move forward.
6. All that said, social media's impact on the world today is real, and it is devastating. The status quo is unsustainable, and these companies need to treat this seriously and make material changes to their platforms, more rapidly than they are currently doing so.

I was named Head of Consumer Product for Twitter in 2015, which meant leading the product teams, in close collaboration with design and engineering, that worked on Twitter for iOS, Android, and the Web—the core consumer experiences that come to mind when you think of “Twitter”.

I left the company on January 18, 2017 and I promised myself I would never again work in ads-based social media.

Almost a year earlier, on April 12, 2016, I had attended a small dinner in San Francisco organized by [Tristan Harris](#), Google's former Design Ethicist, who would go on to found the [Center for Humane Technology](#). Over dinner with [Dave Morin](#) (Facebook's former Head of Platform), [Chris Messina](#) (the inventor of the hashtag), and a small handful of others, Tristan laid out his early understanding of how social networks were stealing control of people's time, without their knowledge, and what the negative impacts of that were.

The concept itself was not novel—how much time people spent on Facebook, Twitter, or Instagram was a frequent topic in popular culture—but the depth of his understanding of it, and his clairvoyance in recognizing the downstream consequences, were both enlightening and terrifying. I was struck by the dinner, and I stepped down from my role in product leadership just a couple months later (I spent the remainder of my time at Twitter aiding the corporate strategy teams).

To be clear, I have been obsessed with data privacy for a long time. When my co-founder Wayne Chang and I sold [Crashlytics](#) to Twitter, in January of 2013, we made sure that the data we brought with us (which already covered hundreds of millions of smartphones worldwide) was explicitly siloed off, and inaccessible to other business, product, or engineering efforts, including advertising. (This was highly atypical at the time, and in stark contrast to Facebook's Onavo acquisition.) Wayne and I apply this same level of care for customer data to this day, at [Digits](#).

A few months after I left Twitter, in the Spring of 2017, Jeff Orlowski reached out to me. He was the director/cinematographer behind the Emmy-winning climate change films [Chasing Ice](#) and [Chasing Coral](#), which I had helped produce. (*N.B. — I am not a producer on The Social Dilemma and have no stake in the project.*) It turns out Jeff had also heard about Tristan's work, and was curious to gut-check what he was learning.

The essence of the problem can be distilled as follows: the ads-based business models that power almost all social networks (both then, and now) implicitly motivate efforts to increase time spent per-user in the product, in order to make more revenue. And the techniques they were using to do so were getting ever-more advanced.

Were we just crazy?

As I discussed this further with Tristan and other friends in tech over the course of 2017, it struck me just how eye-opening and controversial this was, even to those who best

understood the dynamics of how Internet-based social networks operated.

Could Facebook, YouTube, or Twitter really have played any role in Brexit or in the 2016 US Election? Zuckerberg had gone on the record to vehemently deny the possibility.

But reflecting on it with the full benefit of hindsight, the answer is yes. Unquestionably.

Today, it's even more clear. If you ask yourself...

How has the world become so polarized and divisive?

How have lies come to outpace facts on a daily basis?

How have baseless conspiracy theories gained so much in popularity that they are impacting broad populations and mainstream political parties?

...and then rationally walk yourself through the underlying mechanics of the social media business, it all becomes so painfully obvious.

Simply put, these aspects of today's global chaos are each a direct, rational result of trying to make money from ads.

Spoken technically: there are insidious, inescapable, *structural* issues with any consumer product that combines 1) advertising, with 2) user-generated content, and 3) machine learning.

But wait, isn't this just Health & Safety?

It's important to explicitly distinguish *The Social Dilemma* from the essential, long-recognized, and ongoing, health & safety efforts across all the major social platforms today.

Online abuse and harassment is an incredibly serious topic with real-world consequences, and countless individuals both inside and outside the tech industry have risked their careers and reputations over the past decade to drive awareness of these issues and demand changes.

Without at all diminishing those efforts, what is terrifying about *The Social Dilemma* is that it's not actually about health and safety, but something more structural and insidious: the business model itself.

Imagine if everyone on the entire Internet was as friendly and as supportive of each other as the nicest Great British Baking Show contestant... wouldn't that be amazing?

Unfortunately, the devastating consequences of online ads-based business models would still exist. The world today would still become increasingly polarized, and increasingly influenced by lies, just a bit more politely.

Advertising, of course, isn't new. Someone with money to spend and an agenda to promote can pay to get their message in front of an audience. It has been this way for hundreds and hundreds of years.

I choose the TV shows I watch, and the magazines and newspaper articles I read, but I don't choose the ads that come with them. They are picked (for better or worse) by the outlet's publishers and by the sponsors that they work with.

Historically, though, these have all been passive experiences. The content was produced, it was distributed, and it was consumed. And the content, just like the advertisements, was

immutable: predetermined before the consumer arrived, and targeted — if at all — at a broad population demographic. Every consumer enjoyed or suffered the same experience.

Digital advertising started with this same framework in the '90s, with banner ads littering the tops and sidebars of countless online publications. Far more annoying than print ads, to be sure — given their propensity for garish colors and crude GIF animations — but fundamentally identical nonetheless.

Then, everything changed.

What if ads could be selected for display in real-time? What if they could vary based on the person who would experience them? What if they could react to the specific intent of the consumer, in the moment?

Search advertising (invented by Overture, popularized by Google¹) did just this via keywords. Suddenly, ads were radically more relevant. Which, of course, made them radically more effective.

At that time, in the early 2000's, this was largely accepted as an improvement: why would I want to see an ad that I had no interest in? Could ads finally be helpful?!

With the rise of social networks in the mid-2000s, this approach continued. Ads were selected based on the people I followed or the demographic and location information I revealed in my profile.

It seemed reasonable, but there was a critical difference.

With search advertising, the ads I see are a direct result of the actions I take. I consciously decide to search for something and then I see relevant ads as part of the results page. OK, perhaps that is fair. The more searches I do, the more ads I see.

But unlike with search, there's another dimension that can be subconsciously exploited when you combine ads with other media content...

How often have you opened Facebook, YouTube, Twitter, etc., and suddenly 20 minutes pass in the blink of an eye?

This is not an accident. This has been engineered. Why?

Let's imagine *you* are a rationally-behaving social network. Congratulations.

Like any business, you have a fiduciary duty to increase shareholder value: aka drive profits. Since you're ads-based, you have two major levers at hand:

1. You can increase the **relevance** of each ad you show (so advertisers pay you more per ad, because people are more likely to tap on it).
2. Or, you can increase your total number of ad **impressions** (so advertisers pay you more in total, because their ads are shown more times).

Starting with #1, the strive for relevance is the core driver behind these networks' obsession with data collection, and the resulting online privacy protection movement.

The more data you have about me, the more you know about my interests, my strengths, and most importantly my weaknesses. You know what to show me that will get me to watch something, tap something, or buy something — whatever the advertiser wants me to do.

Effectively, ads-based businesses are selling the opportunity to change their users' behavior. Their users' attention and actions are, quite literally, their product.

(This is again why Wayne and I outright refused to share Crashlytics data with Twitter—app developers should not be forced to “sell” their users just to benefit from the functionality of a 3rd-party SDK.)

This part of the story is its own nightmare, but it's a nightmare that the tech and advertising industries generally understand and accept today, for better or worse, and I won't cover it further.

The need to drive ad impressions is far more insidious.

Remember, *you* are now a rationally-behaving social network. When you turn your attention to ad impressions, you have 3 major ways to drive growth:

1. You can get **more monetizable users** (the total number of people you can show ads to).
2. You can **increase ad load** (how many ads are shown to people compared to pieces of normal, organic content).
3. You can **increase usage** (how often and for how long each user engages with your product, so you have the opportunity to show them more ads over the duration).

Monetizable user growth is typically achieved via virality tactics and social psychology. By promoting engagement with new people (“import your contact list!”), notifying them of actions they may be interested in (“you’ve been tagged in a photo!”), and orchestrating feedback loops (“Sally and 5 others liked your post!”), it is conceptually straightforward for social networks to spread through and across real-world communities (“We found 12 people you may know!”).

Ad load is also straightforward and easily tunable, but it can quickly pass a tipping point: most people have a strong distaste for seeing too many ads in too short of a timeframe. In fact, you may notice that towards the end of fiscal quarters, you often see more ads than usual on social networks — an easy way for them to juice the numbers for a few days to ensure revenue targets are hit, then dial it back before too many people get too annoyed.

Finally, **increasing usage** is similar to driving user-base growth. Some social interactions can be tuned to promote reactivating existing users rather than focusing on bringing new people in.

At least, that's what everyone thought...

It turns out you can drive usage without relying on social interactions at all.

Over the course of 2013–14, Facebook famously switched its newsfeed to be algorithmic, where for the first time the content you saw was not purely the result of the people you friended and the order in which they posted things. There were now other factors at play.

Despite widespread threats to “delete Facebook” (surprise! most people didn't), this made sense at first: did I really want to see every single post from that crazy high-school friend who now had too much time on their hands and spent every waking moment on the platform? *(But de-friending them would be so awkward...)*

After all, what is the difference between an ad and a piece of organic content? If Facebook could programmatically select the ads people were seeing, why couldn't its machine-learning algorithms programmatically select the user-generated content people saw as well? In theory, that could result in a much better product experience than pure chronological order.

Their realization of this was profound.

But wait...

Organic content that is “better” in which way? “Better” for whom? How would the algorithm pick which content to show?

Facebook had a fiduciary obligation to drive ad impressions (which drove revenue), so of course that became a major factor in their algorithmic content selection criteria. And sadly, this change was so effective that every other social network had no choice but to mimic it in order to remain competitive.

It turns out that social media usage (time spent per user) is highly influenced not just by social interactions or by the ads you see, but also by the organic, user-generated content you see, *and the order in which you see it.*

For the first time in history, it was possible for a media platform to uniquely specify content, in order, on a person-by-person basis, that had the highest probability of incrementally keeping that person in the product, which meant that person scrolled down further, which meant they saw more ads.

Machine learning algorithms, at a high level, are very easy to understand: give the algorithms a goal state (e.g. drive ad revenue), and they will optimize for it ruthlessly.

The final piece of the puzzle is, sadly, human nature.

We want to belong.

As humans, we tend to not like things we disagree with. We want to be part of a community of shared-interests. We want to be told that we’re right, that we’re not crazy, that there are lots of other people like us.

Otherwise, if we see something we disagree with, our fight-or-flight mentality kicks in. And we close the app.

We also have insatiable appetites for novelty and drama. We want to be shocked, enlightened, fascinated, enraged, or amused, but always in a way that directionally aligns with our existing beliefs.

Polarizing for Profit

In just a couple of years, by the time Brexit and the 2016 election were coming into focus, Facebook’s machine learning algorithms had made some critical “discoveries”:

- As a general rule, people are more likely to seek out and consume content that they already agree with.
- They’re more likely to engage with and share content that is outrageous in some way. (Yes, the algorithms rediscovered what clickbait is.)
- And, once enraged and engaged, people are more likely to keep scrolling for more content along those same lines.

This let Facebook show a lot more, and a lot better-targeted, ads than they were ever able to before. They achieved advertising’s holy grail: they were able to accurately target both content, and ads, to an audience of 1. At a global scale.

The net result is that, purely in the name of ad revenue, social networks are polarizing our world today at a rapid pace.

Every piece of content I see — for hours a day in many people’s cases — is specifically selected to keep me there. To keep me scrolling. To keep me tapping. **To keep me seeing ads.** Because it’s almost always something that I like, or find amusing, or find outrageous — and the algorithms already know that!

So the endgame is sadly clear: the major consumer tech platforms, not just in the US, but globally, now have an implicit fiduciary obligation to reinforce my current views, which only makes me (and everyone else) more dogmatic and more radical along my own lines, no matter which topic the content is about or which direction I lean.

I am being polarized and radicalized for their profit.

Real-world Impact

This isn’t just theory or hyperbole—this impacts literally billions of people every single day —and the real-world effects of this are dramatic:

1. The reality of global warming is undisputed by science and yet a large percentage of the planet has been convinced otherwise, blocking desperately needed regulation and policy changes.
2. The baseless Pizzagate conspiracy theory was “shared roughly 1.4 million times by more than a quarter of a million accounts in its first five weeks of life”² and resulted in an armed gunman storming a DC-area pizzeria.
3. The government of Myanmar systematically, over a period of years, leveraged Facebook to incite hatred towards the Rohingya minority group, to the point where genocide became socially acceptable amongst the majority.³
4. Governments and political groups in Mexico, Brazil, and 70 other countries have been caught intentionally spreading disinformation to their own citizens via social media.⁴
5. Well-known American politicians routinely spread outrageous, brazen lies, knowing that they will go viral amongst their fan bases and that few people will ever take the time to check their veracity. Studies have shown that lies spread 6 times faster than facts.⁵

Never before have media platforms been so incentivized to spread outrage and disinformation, but what can be done?

As with most tragedies of the commons, it’s difficult for any single person to make much difference. Even remaining at Twitter in a position of influence would be a fool’s errand — the problem is with the business model itself, not the execution thereof. Great execution towards business goals is what got Facebook here in the first place.

And this has become a trap: alternative business models, such as charging a subscription for access or charging by follower count/audience size, will almost certainly result in far less total revenue. The algorithms have extracted more dollars from advertisers, on a per-user basis, than the average user’s willingness (or sheer ability) to pay.

In a competitive capitalist market, there is no clear path for any individual to take a stand — any company they succeeded in swaying would immediately be overtaken by less-principled competitors and face certain shareholder revolt.

But by acting together, we can make a difference.

Like has happened in so many other industries, it has come to the point where we have no choice but to force these ads-based consumer tech platforms to internalize their negative externalities.

Controversial, no more.

There's good news: this is no longer all that controversial amongst my friends in tech, and there are now many smart people working on these issues and how to tackle them (both inside and outside these companies).

In sharp contrast to my discussions in 2017, many of the engineers, product managers, and tech execs I've spoken with over the past year now fully understand and agree on the mechanics that are driving this. But they don't yet know what to do about it—so, let's have that discussion together. There may not yet be a silver bullet solution, but here are some questions to ponder:

1. Should there be limits on the amount and types of data about an individual that can be collected and stored in a consumer product?
2. Should there be limits on the maximum number of ads that can be shown to an individual, both in a given session and in a given time interval, whichever lasts longer? (We must restore everyone's control over their own time.)
3. Should personalized advertising even be legal at all? (When driving down the highway, we all see the same billboards, and we can have a shared conversation about them. Yet nobody knows what anyone else sees online—how can you possibly empathize?)

It is past time that the tech industry talks about this topic openly and honestly, and commits to fixing it, because this is just the beginning. Machine learning algorithms are relentless in their pursuit of goal states, and this is not the last psychology trick they will discover — we need to be prepared for the next one.

As Tristan so eloquently puts it, we need not worry about the day when machines overcome human strength (when robots “take over the world”)...

We must worry about the day when machines overcome human weakness, when they can manipulate our behavior to achieve their own goals.

Thanks to the simple, rational pursuit of advertising revenue, that day has already passed.