


2. Digital Harms and Disinformation

Have you seen this shark?

Believe it or not, this is a shark on the freeway in Houston, Texas.
[#HurricaneHarvy](#)



11:00 PM - 27 Aug 2017

85,254 Retweets 143,836 Likes

6.9K 85K 144K

2

TL;DR: Digital Harms	2
Digital Harms	2
Hate Speech	3
Misinformation, Malinformation, Disinformation	3
Motivations	4
Money	4
GeoPolitics	6
Politics and Power	7
Business	7
Attention and Fun	8
Mechanisms	9
Targets	9
Channels	10
Values	11

Lessons from Other Disciplines	11
AdTech: Microtargeting	12
Data Science: why disinformation is everywhere now	12
Infosec: How Disinformation Might Evolve	13
Further Reading	13

TL;DR: Digital Harms

- *Disinformation is deliberate promotion of false, misleading or misattributed information.*
- *You're not here to stop debate, however odious it is - you're here to reduce online harms.*
- *Disinformation motivations include geopolitics, money, politics, power, and attention.*

Digital Harms

Disinformation is a [digital harm](#), alongside ransomware, cyberbullying, hate speech, and spam. Effects include:

- **Physical.** Bodily injury, damage to physical assets (hardware, infrastructure).
- **Psychological.** Depression, anxiety from cyber bullying, cyber stalking etc
- **Economic.** Financial loss, e.g. from data breach, cybercrime etc
- **Reputational.** Organization's loss of consumers, individual's disruption of personal life, country's damaged trade negotiations.
- **Cultural.** Increase in social disruption, creating [real-world violence](#).
- **Political.** Disruption in political process, lost government services from internet shutdown, botnets influencing votes

Always look for the harms, and the motivations for those harms.

Hate Speech

"Hate speech, speech or expression that denigrates a person or persons on the basis of (alleged) membership in a social group identified by attributes such as race, ethnicity, gender, sexual orientation, religion, age, physical or mental disability, and others." - [Britannica.com](https://www.britannica.com)

Misinformation, Malinformation, Disinformation

Misinformation is false content: untruths in text, images, videos. That false content might be unintentional, or might be part of a coordinated disinformation effort. Classic examples of misinformation include rumours about Covid-19 folk cures, and stories about sharks swimming in flooded subways.

Malinformation is information that's true, private, and posted publicly to cause harm. A classic example of malinformation is a political "hack and leak", where private information is stolen then shared online.

There are many definitions of "disinformation": pick one that works for you and your practical work. The CogSecCollab definition of disinformation is *"deliberate promotion of false, misleading or misattributed information. We focus on the creation, propagation and consumption of disinformation online. We are especially interested in disinformation designed to change beliefs or emotions in a large number of people"*¹. That allows us to talk about:

- intentionality ("deliberate promotion"),
- non-false information ("misleading or mis-attributed"),
- goals ("designed to change beliefs or emotions in a large number of people") and
- mechanisms ("focus on creation, propagation, consumption of misinformation online").

¹ From the Credibility Coalition's MisinfoSec Working Group

The Global Disinformation Index uses signals of intent, e.g. do these sites contain hate speech, are they targeting specific groups etc. This changes the focus from looking for something subjective, intangible and subject to bias (e.g. political sites are very difficult to flag as misinformation/not), to more objective tagging.

Disinformation isn't misinformation². Disinformation is intentional, and its falsehood isn't always in its content: many successful disinformation campaigns use factual information, or a mix of factual information and misinformation³. Disinformation's falsehoods are often contextual: how information is labelled, who it comes from, its apparent popularity through amplification, and groups set up as channels for it.

Motivations

People produce disinformation for attention, power, money, and political or geo-political gain.

Not all misinformation is harmful. The social internet is driven by community: online discussion includes rumour, opinion, conspiracy theories, protests, extremists and combinations of these. These might be distasteful, but not disinformation. Look for harms, and the coordinated inauthentic activities that potentially cause them.

Money

Money is a popular motive for creating disinformation, which drives users to websites, and sales. Ways to make money from disinformation include:

² See [Claire Wardle's work](#) on the differences between misinformation and disinformation

³ One report suggests the most effective ratio is 90% true to 10% misinformation content.

- get people to look at a website, click on something, or do something like fill out a form; this can produce advertising revenue, and [personal data that can be sold](#). Metrics for this include cpm: \$ for every thousand eyeballs, cpc: \$ for every click - a lot higher than cpm because it's a lot rarer, and cpa - \$ for an action, and usually much rarer.
- sell merchandise - t-shirts, books, videos, 'cures'; or services - speaking at events
- sell or rent accounts, including individual accounts, and botnets - which are still cheap
- sell disinformation services: spam farms for disinfo, or creating deep fakes - currently at about \$2 per hour and 5-6 hours per fake

People making money online usually need a place to do it in - a web domain, or some other place that they can push people towards. Online applications or platforms have several ways to make money:

- One-off payments (e.g. buying a t-shirt from an online vendor)
- Commissions (e.g. Amazon percentages on marketplace)
- Subscriptions (e.g. Spotify premium, AWS, New York Times etc)
- Online advertising (e.g. selling advert views, clicks, actions on your webpages or videos)

In 2016, the "Macedonian Teens" and other website builders discovered that political anger, outrage, and fear could attract people to view their sites. In 2020, the fear was of both disease and its cures: websites ranged from anti-vaccination and anti-mask information, to sites selling alternative 'cures' for Covid19. Moneymakers often hitch a free ride on disinformation narratives and groups created for other purposes. Affiliate marketing is also popular - usually to its own network of sketchy sites.

GeoPolitics

Countries use disinformation to change external opinions of themselves, their actions, and the state of areas they have interests in, and to weaken the population and environments of their potential opponents. Disinformation is cheaper than conventional warfare, with very few current downsides for a country willing to use it, can be outsourced to small teams and individuals outside the country using or the subject of it, and can be designed to continue in the target country long after the creating team has moved on.

Nation states, and some non-state actors including terrorist and transnational crime groups, influence each other through “instruments of national power”, usually referred to as the DIME model:

- Diplomatic: organizing coalitions and alliances, which may include states and non-state entities, as partners, allies, surrogates, and/or proxies
- Informational: using information to further their causes and undermine those of other countries and allegiances
- Military: compel an adversary, or resist external compulsion, through the threat or application of force
- Economic: An economy with free access to global markets and resources is a fundamental engine of the general welfare, the enabler of a strong national defense.

These instruments of national power are how countries maintain their sovereignty and influence other nations. Informational instruments include public affairs, public diplomacy, communications resources, spokespersons, timing, and media.

Democracies require common knowledge (who the rulers are, legitimacy of the rulers, how government works), draw on contested political knowledge to solve problems, and are

vulnerable to attacks on common political knowledge. Autocracies actively suppress common political knowledge, benefit from contested political knowledge and are vulnerable to attacks on the monopoly of common political knowledge.

Politics and Power



2020 social media message

Politicians, organisations, and extremist groups use disinformation to affect people in their own country. Geopolitical actors also create, subvert, or hijack political and activist groups. Most political disinformation narratives are designed to reject 'out-groups', and increase the coherence of 'in-groups' to create strong groups of followers.

Business

There's a business equivalent to the DIME model⁴:

- Business deals and strategic partnerships
- PR and advertising
- Mergers and acquisitions
- R&D and capital investments

⁴ See Pablo Breuer and David Perlman's 2018 Black Hat talk for more

All of these can be attacked using disinformation campaigns.

Attention and Fun

Attention-seeking misinformation is usually smaller-scale, short-term and created by individuals. Examples in disasters include realistic calls for help from individuals seeking attention - e.g. the fake tweet “I’m stuck under a building with my child” during the Chile 2010 earthquake, but might also be nationstates testing disinformation mechanisms⁵.

Attention-seeking misinformation usually gets lost in the noise, unless it's flooding an important hashtag, area, or group - the social media equivalent of a DDOS, e.g. blocking a crisis hashtag that data responders are social listening on, looking for information they can add to a disaster situation picture and/or route to responders.

Misinformation-for-fun going viral has a long history. One example is the disaster shark: in almost every natural disaster in the last decade, someone has posted a picture of the same shark as “sharks in the street”, “sharks in the subway” etc, and pushed it to go viral.

Generally, misinformation for LOLs isn't an issue, unless it's satire and conspiracies being used as a gateway into more worrying narratives and groups.

Responses to attention-seeking and fun-based misinformation include triple-verifying, e.g. don't post any information until it's seen and checked in 3 sources; to reach out and ask the poster to remove the misinformation - including the reason why; and to push back with a counter-message - gentle humour can be good. Typically, people posting misinformation for fun are amenable to helping counter any ill effects from it, and are less likely to engage in counter-counter games.

⁵ see Kate Starbird's analysis of 2010 BP Oil Spill “tsunami warning” tweets on #oilspill

Mechanisms

Users and groups are influenced online in many ways: disciplines that do this include user experience, marketing and adtech, political campaigns, and psyops. The mechanisms used by them can be adapted for disinformation.

Targets

Disinformation uses people the way that malware uses PCs. Sometimes people, and clusters of people (communities, nations etc) are the endpoints, and sometimes they're channels (e.g. influencers, media) to reach more people, to spread narratives, create confusion or increase community fragmentation and distrust.

Countries target other countries' populations, to weaken those countries through peoples' distrust of each other, their governance systems, and officers of governance, and persuading them to act in ways counter to a strong nation state. Countries also target their own populations, e.g. attacking the credibility of non-ruling parties, voting systems or minorities to stay in power. Successful gambits include increasing distrust between internal groups, often by targeting disinformation campaigns at one or all of the groups around a divisive debate.

Fraudsters target anyone who will give them money. Often this is as simple as building campaigns around getting eyeballs onto a sales site (or just a website: eyeballs and clicks are worth advertising money), by piggybacking on divisive or emotionally-charged conspiracy narratives like Covid5G.

Groups target groups, organisations, individuals. Hospitals were directly targeted as part of the "covid isn't real" narrative. Disinformation campaigns have also targeted individuals including Bill Gates and Anthony Fauci.

Some companies have used disinformation to alter rivals' prospects, but commercial disinformation appears to be generally spam and marketing companies pivoting to a new line of business, Disinformation As A Service.

Channels

- 
- Social networks (examples include MySpace, Facebook, and LinkedIn)
 - Micro-blogging websites (examples include twitter and StumbleUpon)
 - Blogging and Forums websites (examples include WordPress, tumblr, and LIVEJOURNAL)
 - Pictures and Video-Sharing websites (examples include YouTube, flickr, and Flikster)
 - Music websites (examples include Pandora, lost.fm, and iLike)
 - Online Commerce websites (examples include eBay, amazon.com, and Epinions)
 - Dating Network websites (examples include match.com, eHarmony, and chemistry.com)
 - Geo Social Network websites (examples include foursquare, urbanspoon, and tripadvisor)
 - News and Media websites (example include the LA Times, CNN, and New York Times)

Figure: Chris Burgess, types of online interactions

Disinformation can appear anywhere that people share content with each other online.

A large proportion of the world's population can now broadcast instantly to almost everyone else online through pcs and phone apps. This "user-generated content" includes messages, posts, comments, videos, articles, websites and webpages. The channels available to do this include social media, messaging apps, their own websites, and comments on specialised sites including shopping, music, dating, games, news, entertainment, and research apps and websites.

The internet isn't the only system carrying disinformation: we still have word of mouth, merchandising and traditional media like radio, television and newspapers, which are increasingly part of larger disinformation systems.

Values

People use social media for many things, including sharing information and talking to each other. Information flow is well studied, but it isn't the only commodity available where large numbers of people congregate. Human qualities that can be 'hacked' by disinformation campaigns include:

- Viewpoints. The stances that people take on issues, e.g. who shot down MH17.
- Emotions. Much disinformation is designed to change peoples' sentiments towards a viewpoint, group, individual etc.
- Belonging. Finding your community is much easier with billions of people online.
- Purpose.
- Connections. Visibility builds relationships - whether this is with online dates, friends of friends or brands, products and influencers.
- Trust.
- Convening power. The ability to create events and build offline communities.

Lessons from Other Disciplines

When we talk about security going back to thinking about the combination of physical, cyber and cognitive, people sometimes ask why now? Why, apart from the obvious weekly flurries of misinformation incidents, are we talking about cognitive security now? And what's likely to happen next? Disciplines including Data Science, Marketing, and Cybersecurity can offer us some clues.

AdTech: Microtargeting

Much of the online advertising industry is geared to optimising the high-speed auction between advertisers and online property owners (websites, videos, TV, internet-connected billboards etc), to get advertisers coverage whilst optimising the property owners' profits. What they're selling is users' views and actions. And what they optimise on is demographics (for individuals) and Know Your Customer (for businesses).

The difference between online marketing and disinformation campaigns is in intent. It's why we talk about "coordinated inauthentic activity", which focuses on the scale, the behaviour (you can do a good disinformation campaign with true content - e.g. almost any african-american focussed one) and the intent to deceive - where that intent is usually to do some form of harm, whether it's to shape a geopolitical narrative away from the country it's targeted at, or to widen divisions across society.

Most disinformation campaigns look like marketing campaigns because that's where their roots are. The Internet Research Agency was a marketing team that was asked to do a side gig; and many of the new disinformation farms in e.g. the Philippines are repurposed marketing agencies or spam factories.

Data Science: why disinformation is everywhere now

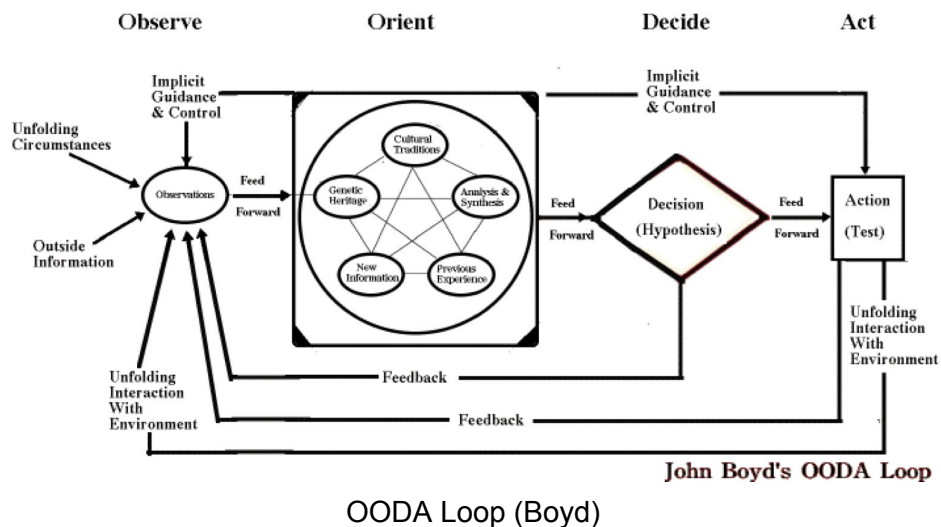
Data Science includes the three Vs of big data: Volume, Velocity, Variety. Adapted for disinformation, these are:

- Volume: Online volumes are high enough that brands and data scientists can spend their days doing social media analysis, looking at cliques, message spread, adaption and reach.

- **Velocity:** Online data is coming in so fast that an incident manager can do AB-testing on humans in real time, adapting messages and other parts of each incident to fit the environment and head towards incident goals faster, and more efficiently. Ideally that adaptation is much faster than any response, which fits the classic definition of “getting inside the other guy’s OODA loop”.
- **Variety:** The internet has a lot of text data floating around it, but its variety isn’t just in all the different platforms and data formats needed to scrape or inject into it — it’s also in the types of information being carried. Everyone and their grandmother is online now, and the (sniffable, actionable and adjustable) data flows include emotions, relationships, group sentiment, market sentiment, and group cohesion markers.

There’s also a fourth V, Veracity, that includes disinformation.

Intelligence: OODA loops and intelligence cycles



Boyd's OODA Loop is a way to look at the process from collecting information about the world, through building a "situation picture" that models what you think might be causing

those observations, deciding how to act in that situation, then acting. It's also a useful lens for thinking about disinformation response in terms of counter-actions, competing decision loops and resources.

Infosec: How Disinformation Might Evolve

Disinformation defence is moving on a similar arc to the one that Cliff Stoll's "[The Cuckoo's Egg](#)" and Mike van Putte's "[Walking Wounded](#)" describe as the evolution of cybersecurity.

2016 disinformation was roughly equivalent to the start of The Cuckoo's Egg, where Stoll starts noticing there's a problem in his systems, and tracks hackers moving through them. Disinformation is a bit further now. There is now a market for disinformation as a service. Disinformation response is also a market, but it's one with several layers to it, just as the existing cybersecurity market has specialists and sizes and layers.

Further Reading

Recent information operations, disinformation and propaganda history:

- Thomas Rid, "[Active Measures](#)"
- PW Singer, Emerson Brooking "[Like War](#)"
- Zeynep Tufekci, "Twitter and Tear Gas" ([free version](#))
- "[Verification handbook](#)", specifically the chapter on [investigative reporting](#)

Understanding information security:

- [Rent-a-troll: Researchers pit disinformation farmers against each other](#)
- [Market Sentiment](#)

Internet history:

The BigBook of Disinformation Defence v2.0

- [An Internet History Timeline: From the 1960s to Now](#)
- <https://www.slideshare.net/debbylatina/internet-history-190741201>
- We Are Social: [Global digital report 2019](#) - internet size

Abuses and counters

- [I stumbled across a huge Airbnb scam that's taking over London](#)
- Ethan Zuckerman course, "[Fixing Social Media](#)"
- [There are no sharks swimming on a freeway in Houston](#)
- Kate Starbird, [Tracing Disinformation Trajectories from the 2010 Deepwater Horizon Oil Spill](#), 2016

Human vulnerabilities:

- Jonathan Haidt "why it feels like everything is going haywire"
- [Demand for Deceit: Why Do People Consume and Share Disinformation? – Power 3.0: Understanding Modern Authoritarian Influence](#)

History of geopolitical influence:

- [Final Report on the Bulgarian Broadcasting Station New Europe, \(Research Unit X.2\)](#)
- [Morale Operations FM](#)
- [Unrestricted Warfare](#)
- <https://www.psywar.org/content/sibsLecture>
- [Russian Political War | Moving Beyond the Hybrid](#)

Geopolitical disinformation

- H. Farrell and B. Schneier "Defending Democratic Mechanisms and Institutions against Information Attacks" Shneier on Security, 2019

- H. Farrell & B. Schneier “Common-Knowledge Attacks on Democracy” Berkman Klein Center for Internet and Society. Harvard University. October, 2018
- S.C. Wooley & P.N Howard (eds) Computational Propaganda. Oxford. 2019