# Chapter 1 Disinformation


**TL;DR: Looking for Disinformation** **1**

**Defining Disinformation** **2**

Good introductions to disinformation 3

**Online Influence and its abuses** **3**

What People Do Online 3

With people comes value 4

With value comes targetting 4

With value comes abuse 5

Disinformation as a Digital Harm 5

Reading 6

**Disinformation from the Creators' POV: Intent** **6**

**Where disinformation comes from** **6**

Disinformation for Money 7

Disinformation for GeoPolitics 8

Disinformation for Politics 9

Disinformation for Power 9

Disinformation for Business 9

Disinformation for Lols and Attention 10

Satire and Conspiracies 10

Reading 11

**Disinformation from the Defender POV** **11**

What disinformation targets 11

Big, Fast, Weird: why disinformation is getting harder to track 12

Further Reading 13


## TL;DR: Looking for Disinformation

You're looking for deliberate promotion of false, misleading or misattributed information in your team's specific area of interest. This might be motivated by geopolitics (countries interfering with the beliefs and sentiments of each others' populations), money (selling webpage views, clicks, t-shirts, 'cures' etc), politics, power (e.g. adding people to extremist 'in-groups'), or LOLs. You're not here to stop debate, however odious it is - you are here to help reduce online harms.

# Defining Disinformation

We look at disinformation as an information security threat. It helps to define how we see the threat, its sources and its manifestations.

The short answer on defining disinformation is “don’t get hung up on definitions". There are many definitions of disinformation, misinformation, malinformation, propaganda, influence etc., and standards working groups dedicated to defining the differences between them. If you’re working on practical disinformation response, try not to get sucked into that very large rabbithole: pick a working definition for yourself, work out what matters to your practical work, and focus on that.

For instance, the one we're using here comes from the Credibility Coalition's Misinfosec Working Group: "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".

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”).

We are countering deliberate creation and propagation of false information online - whether that falsehood is in the information itself, who it purports to come from, the groups set up to output it, etc. We care a lot when it’s designed to change beliefs or emotions in large numbers of people.

Within another organisation, we switched from trying to define misinformation in websites, to looking at signals of intent, e.g. did these sites contain hate speech, were they targetting specific groups etc. That moved the definitions from 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 the same thing as misinformation.

- Misinformation is false content: untruths in text, faked images etc; and those might be unintentional, or not be part of a coordinated effort. Disinformation is false. It’s intentional. It’s at scale. And the falsehood might not be in the content - the content, or the original poster’s intentions, might be clean, but the reuse, or amplification etc might be designed to create harm.
- A third category is Malinformation: information that's true, but usually private, and posted online to cause harm.

A good place to start if you want to dig into these definitions is Clare Wardle's 2017 work on Information Disorder.

Disinformation and Malinformation are examples of online harms, alongside things like ransomware, cyberbullying, etc. Always look for the harms, and the motivations for those harms.

### Good introductions to disinformation

If you say you're working on disinformation, people around you will often quietly ask how they can help, and where they can get more information about it. Good introductions that you can show your mum and other people who ask include:

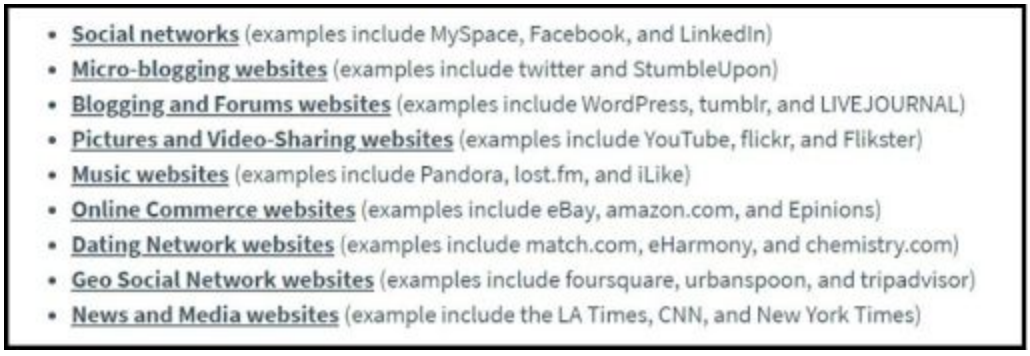
- The War on Pineapple: Understanding Foreign Interference in 5 Steps
- Bad News Game
- The Dark(er) Side of Media: Crash Course Media Literacy #10
- Web Literacy for Student Fact-Checkers – Simple Book Production

Although that doesn't cover everything we do, those references between them give a good introduction to what we're dealing with, and some of the things that everyone can do to help mitigate them.

## Online Influence and its abuses

Let’s look at the range of ways users and groups are influenced online (and offline via online means) - user experience, marketing and adtech, online political campaigns, astroturfing, online psyops, disinformation campaigns.

### What People Do Online



- **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

The internet has changed a lot since the early days of ARPANET, JANET and bulletin boards. People still do the same things - sharing information and talking to each other - but the ability to do that isn't limited to the techies and companies who could pay for website designs, and the volume, variety and velocity of information and the people and organisations receiving it has increased to encompass (through localisation, phone apps etc) a large proportion of the world's population. Anyone can broadcast to almost anyone else almost instantly over a large number of specialised (shopping, music, dating, games, news, entertainment, research, etc) and general (social media, blogs, new websites etc) sites, through user-generated content like messages, posts and comments, and commercial content like videos, articles and pages.

## With people comes value

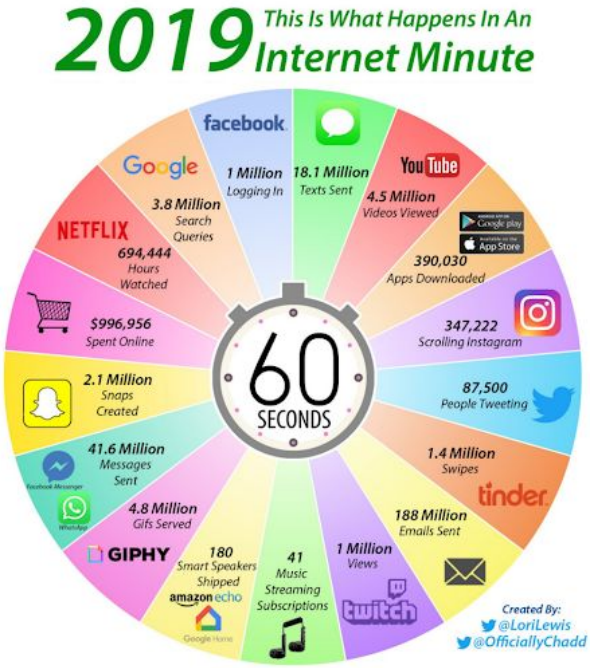



Figure: Internet Minute, 2019

If you create an online application or platform, there are 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)

Money isn't the only commodity available where large numbers of people congregate. Other values include:

- Viewpoints. The stances that people take on issues like, for instance, who downed MH17.
- Belonging. Finding your community is much easier with billions of people online.
- Convening power.  Sites like Eventbrite and Meetup help users build communities offline.
- Connections.  Visibility builds relationships - whether this is with online dates, friends of friends or brands, products and influencers.
- Information.

## With value comes targetting

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).

- Demographics: know your targets
- B2B: Know Your Customer

## With value comes abuse

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 targetted 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; many of the new disinformation farms in e.g. the Philippines are repurposed spam factories etc.

## Disinformation as a Digital Harm

Disinformation is just one form of online abuse, amongst hate speech, spam, online bullying etc. These are often known collectively as "digital harms".

- Physical harm: e.g. bodily injury, damage to physical assets (hardware, infrastructure, etc).
- Psychological harm: e.g. depression, anxiety from cyber bullying, cyber stalking etc
- Economic harm: financial loss, e.g. from data breach, cybercrime etc
- Reputational harm: e.g. Organization: loss of consumers; Individual: disruption of personal life; Country: damaged trade negotiations.
- Cultural harm: increase in social disruption, e.g. [misinformation](https://www.washingtonpost.com/politics/2020/02/21/how-misinformation-whatsapp-led-deathly-mob-lynching-india/) creating real-world violence.
- Political harm: e.g. disruption in political process, government services from e.g. internet shutdown, botnets influencing votes

When we assess whether to respond to a disinformation incident, we use a set of criteria that include an estimate of the risk, defined in terms of the potential harm caused by the incident without response, for example in the League, we're specifically looking for, and trying to reduce, medical harms (another criterion is whether other teams are already responding to reduce the incident's potential harm). The use of the harms framework above is important because in 2020, social media companies, politicians etc shifted their definition of 'bad' on social media from immediate calls for violence to the idea of digital harms where the effects might be delayed.

### Reading

Internet history:

- 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

## Disinformation from the Creators' POV: Intent

We track disinformation incidents and persistent threats. Understanding what disinformation creators do can be improved by first understanding why they do it, and seeing how they might optimise against those goals.

## Where disinformation comes from

The short answer is that people produce disinformation for attention, power, money and political or geopolitical gain.

- Using disinformation for geopolitical gain is very much in the headlines. Countries use it to change 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 done right will continue in the target country long after the creating team has moved on.
- Internal groups and organisations also use disinformation to gain power, often by emphasising ingroup/outgroup narratives to create strong groups of followers.
- Money is a popular motive,and even with other types of disinformation, there are often hucksters riding narratives and groups to make profits.
- Attention-seeking with online disinformation has been around a long time (e.g. the sharks in the street that appear online for most natural disasters, and satire and other LOLs); usually it's smaller-scale and driven short-term by individuals. Mostly, unless it's DDOSing a hashtag or area that's important (e.g. a crisis reporting hashtag) this type of disinformation gets lost in the noise.
- The social internet is driven by community: online discussion contains a lot of misinformation, including rumour, opinion, conspiracy theories, protests, extremists and combinations of them. This is humans being humans. We're not here to stop debate: disinformation tracking is about finding the coordinated inauthentic activities that potentially do harm.

The big idea here is that the Internet makes everyone a nation-state actor now.

## Disinformation for Money


Covid19-related sales included t-shirts, cures, blood, views, clicks

Money: It's grifting. Ways to make money from disinformation:

- get people to look at your website (cpm: $ for every thousand eyeballs), or click on something (cpc: $ for every click - a lot higher than cpm because it's a lot rarer), or do something like fill out a form (cpa - much much rarer usually); inadvertently give you data that you can sell
- sell merchandise like t-shirts, videos, 'cures'; sell services (e.g. the covid5g guy selling books and on a speaking tour);
- sell disinformation services (e.g. like spam farms, but for disinfo, or creating deep fakes - but at about $2 per hour and 5-6 hours per fake that's not making much right now);
- sell or rent accounts (e.g. botnets - again, still relatively cheap)

In 2016, it was the "Macedonian Teens" (in practice, not teens, and not all from Macedonia, but there were some Macedonian villages that were centres for disinformation production), and US-native peeps who discovered that political outrage on the right side of the spectrum got more clicks. Basically: anger and fear sell

Now, there are a lot of antivax sites selling 'alternative cures'. Also good bit of affiliate marketing - usually to its own network of sketchy sites.

## Disinformation for GeoPolitics

Since 1648 (the end of the 30 Years War, where over 8 million people died), modern international discourse between nations has been based on Westphalian Sovereignty. This includes the principles:

- Each nation has sovereignty over its own territory and domestic affairs
- No nation should interfere in another country's domestic affairs
- Each state is equal under international law

Nation states influence each other through the instruments of national power; levers they can pull on to influence other nations. These are resources available in pursuit of national objectives, usually referred to as the DIME model \[74\]:

- Diplomatic: Diplomacy is a principal means of organizing coalitions and alliances, which may include states and non-state entities, as partners, allies, surrogates, and/or proxies I
- Informational: The concept of information as an instrument of national power extends to non-state actors—such as terrorists and transnational criminal groups—that are using information to further their causes and undermine those of the USG and our allies.
- Military: Fundamentally, the military instrument is coercive in nature, to include the integral aspect of military capability that opposes external coercion. Coercion generates effects through the application of force (to include the threat of force) to compel an adversary or prevent our being compelled. The military has various capabilities that are useful in non-conflict situations (such as in foreign relief).
- 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. In the international arena, the Department of the Treasury works with other USG agencies, the governments of other nations, and the international financial institutions to encourage economic growth, raise standards of living, and predict and prevent, to the extent possible, economic and financial crises.

These instruments of national power are how countries maintain their sovereignty and influence other nations.

In practice these instruments overlap. In particular, informational instruments include public affairs, public diplomacy, communications resources, spokespersons, timing and media. For a long time, the ability to reach mass audiences belonged to the nation-state (e.g. in the USA via broadcast licensing through ABC, CBS and NBC). Now, however, control of informational instruments has been allowed to devolve to large technology companies who have been blissfully complacent and complicit in facilitating access to the public for information operators at a fraction of what it would have cost them by other means.

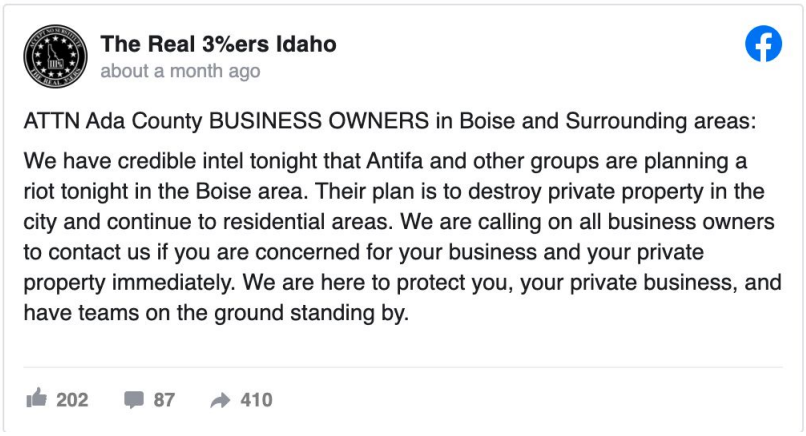
Democracies and autocracies appear to have different vulnerabilities to information threats [Farrell19] [Farrell18] [Wooley19]. Democracies require common knowledge (who the rulers are, legitimacy of the rulers, how government works), draw on contested political knowledge to solve problems, and are vulnerable 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.

## Disinformation for Politics

## Disinformation for Power

SIGN UP




2020 social media message

There are groups who use disinformation for power, but who are not (overtly) part of political parties or geopolitical actions, although they're often directly or indirectly attached to political groups or created/ subverted/ hijacked by geopolitical actors.

Many of these are far-right-wing groups internal to the countries they operate in, but disinformation is always a tempting tool for other activist groups.

There are also groups that use disinformation campaigns to create and use other forms of power. As an example, many of the tactics used by modern power-disinformation groups can be traced back to anti-feminist groups and actions like \#gamergate.

## Disinformation for Business

There's a business equivalent to the DIME model (Breuer&Perlman):

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

All of these things can be attacked using disinformation campaigns.

## Disinformation for Lols and Attention


The disaster shark. Every natural disaster, the same shark

Misinformation-for-fun going viral online has a long history. One classic example of this is the disaster shark pictures: 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. Crisismappers see it and sigh, then ask the original poster to remove it please because it’s messing up the online response (crisismappers are typically social listening on disaster-related hashtags, looking for information they can add to a disaster situation picture and/or route to responders). 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.

T-shirt of fake tweet sent during Chile 2010 earthquake

Misinformation for attention also has a long history. Crisismappers have worked since 2010 to remove both for-profit (Ugg advertisements) and well-meaning (“thoughts and prayers”) spam from their feeds, but also to handle deliberate injections of what looks like real disaster-related information, both from individuals seeking attention (e.g the Chile t-shirt tweet above), and from nationstates testing disinformation mechanisms (see Kate Starbird’s analysis of the 2010 BP Oil Spill “tsunami warning” tweets on \#oilspill). Generally, crisismappers triple-verify, e.g. don’t post any information until we’ve received it from 3 sources and checked each of them out; for misinformation the process is to gently push back with a message (gentle humour can be good), and to reach out and ask the poster to remove it from social media.

## Satire and Conspiracies

Generally, we ignore disinformation for LOLs unless it's flooding an important hashtag or group (the social media equivalent of a DDOS). The only caveat on that is that satire and conspiracies are sometimes used as a gateway into more worrying narratives and groups, and might merit attention for that.

## Reading

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)
- https://www.psywar.org/articles
- Morale Operations FM
- “Oss morale operations”
- Unrestricted_Warfare
- https://www.psywar.org/content/sibsLecture
- Russian Political War | Moving Beyond the Hybrid
- Also check https://www.psywar.org/articles

Geopolitical disinformation

- [Farrell19] H. Farrell and B. Schneier “Defending Democratic Mechanisms and Institutions against Information Attacks” Shneier on Security, 2019
- [Farrell18] H. Farrell & B. Schneier “Common-Knowledge Attacks on Democracy” Berkman Klein Center for Internet and Society. Harvard University. October, 2018
- [Wooley19] S.C. Wooley & P.N Howard (eds) Computational Propaganda. Oxford. 2019

Country-specific datasets

- EuVsDisinfo database. Database of pro-Kremlin disinformation https://euvsdisinfo.eu/disinformation-cases/. Ordered by date, narrative, outlets and countries, with summary and disproof. Described in https://euvsdisinfo.eu/old-wine-new-bottles-6500-disinformation-cases-later/. Publicly accessible, no API.
- Facebook GRU dataset provided to SSCI. Not publicly available; described in “Potemkin Pages & Personas”
  Omelas https://www.omelas.io/ has a live feed, multiple countries (Russia, China etc) but I don't think they've gone public with their dashboard yet - can ask for email summaries
- Russia analysis: KremlinWatch does analysis on Russia-EU ops https://www.kremlinwatch.eu/#welcome ; CEPA is more high-level http://infowar.cepa.org/This-week-in-infowar.
  If you’re looking for non-Russia, you’re basically looking at specialists.

# Disinformation from the Defender POV

## What 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 sometimes target other countries to weaken them by helping populations distrust each other and their systems and officers of governance, and 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.

There has been some directly targeted disinformation. Individuals (BillGates, Fauci) have had targeted disinformation campaigns around them; some campaigns directly targeted hospitals as part of the "covid isn't real" narrative, and some companies have used disinformation to alter rivals' prospects. Some hybrid infosec/disinformation attacks using deep faked voices also exist but are still relatively rare compared to e.g. ransomware. Commercial disinformation appears at the moment to be generally spam and marketing companies pivoting to disinformation as a service as a new line of business.

## Big, Fast, Weird: why disinformation is getting harder to track

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?

One answer is the three Vs of big data: volume, velocity, variety (the fourth V, veracity, is kinda the point of disinformation, so we're leaving it out of this discussion).

- 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. We're way past the Internet 1.0 days of someone

posting the sports scores online and a bunch of hackers lurking on bulletin boards: now everyone and their grandmother is here, and the (sniffable, actionable and adjustable) data flows include emotions, relationships, group sentiment (anyone thinking about market sentiment should be at least a little worried by now) and group cohesion markers.
- Volume: There's a lot of it — 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: And it's coming in fast: 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, more efficiently etc. Ideally that adaptation is much faster than any response, which fits the classic definition of "getting inside the other guy's OODA loop".

NB The internet isn't the only system carrying these things: we still have traditional media like radio, television and newspapers, but they're each increasingly part of these larger connected systems.

Another common question is "so what happens next". One answer is to point people at two books: Cliff Stoll's "The Cuckoo's Egg" and Mike van Putte's "Walking Wounded" — both excellent books about the evolution of the cybersecurity industry (and not just because great friends feature in them), and say we're at the start of The Cuckoo's Egg, where Stoll starts noticing there's a problem in the systems and tracking the hackers through them.

We're getting a bit further through that book now. In America, if someone sees a threat, someone else makes a market out of it. Cuddle-an-alligator — tick. Scorpion lollipops in the supermarket — yep. Disinformation as a service / disinformation response as a service — also in the works, as predicted for a few years now.

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. One of the reasons for working on disinformation threat intelligence is to help encourage that market to grow.

## Further Reading

If you really want to get into how we got here, the history of information operations, what disinformation and propaganda are etc, these books were recommended by the team:

- [SJ's 2018 book stack - dated, but some good classics in here](http://overcognition.com/2018/12/10/misinformation-readings/ )
- Thomas Rid's "[Active Measures](https://us.macmillan.com/books/9780374287269 )"
- PW Singer and Emerson Brooking's "[Like War](https://www.likewarbook.com/ )"

- Zeynap Tufeki's "[Twitter and Tear Gas](https://www.twitterandteargas.org/downloads/twitter-and-tear-gas-by-zeynep-tufekci.pdf )" (free version)
- Verification handbook: http://verificationhandbook.com/, specifically the chapter on investigative reporting http://verificationhandbook.com/book2/chapter1.php

Understanding infosec:

- The Cuckoo's Egg
- Walking Wounded
- Rent-a-troll: Researchers pit disinformation farmers against each other
- Market Sentiment