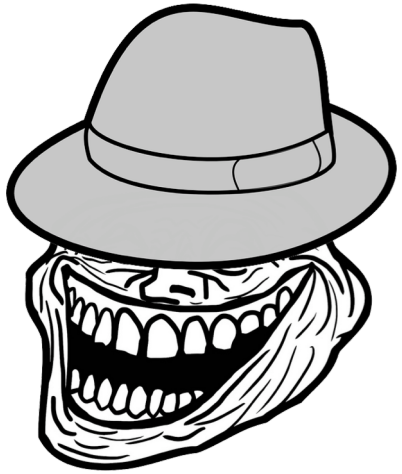


# Beyond the Lulz:



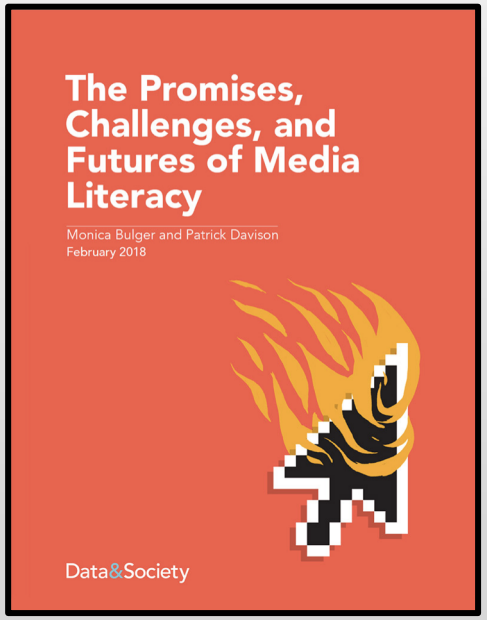
Black Hat Trolling, White Hat Trolling, and Hacking the Attention Landscape

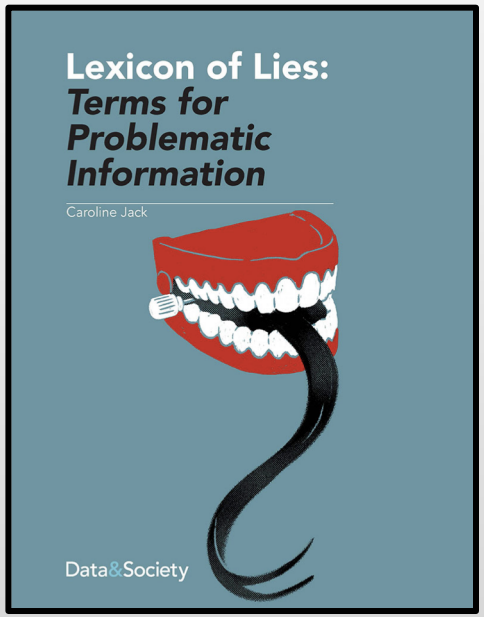


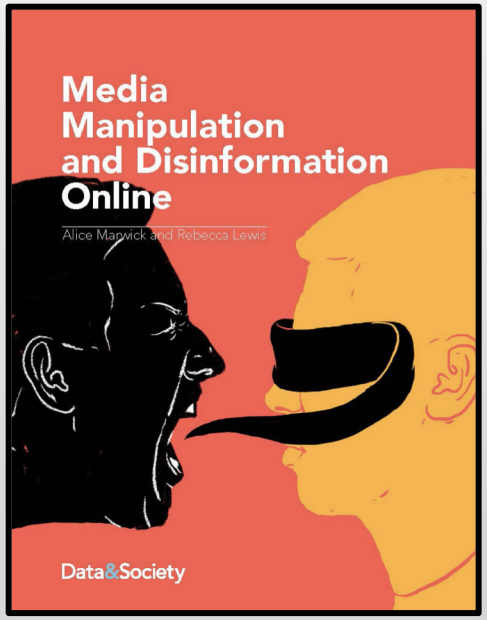


# CONTEXT


Dead Reckoning
Navigating Content Moderation After "Fake News"
February 2018
Robyn Caplan, Lauren Hanson, and Joan Donovan
Data&Society

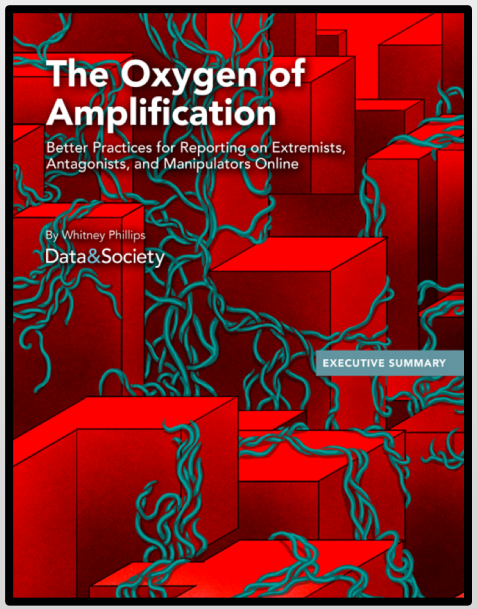

The Promises, Challenges, and Futures of Media Literacy
Monica Bulger and Patrick Davison
February 2018
Data&Society


Lexicon of Lies: Terms for Problematic Information
Caroline Jack
Data&Society


Media Manipulation and Disinformation Online
Alice Marwick and Rebecca Lewis
Data&Society

# CONTEXT


Searching for Alternative Facts:
Analyzing Scriptural Inference in Conservative News Practices
Francesca Tripodi
Data&Society


The Oxygen of Amplification
Better Practices for Reporting on Extremists, Antagonists, and Manipulators Online
By Whitney Phillips
Data&Society
EXECUTIVE SUMMARY

**SECTION 1**





# HACKING and TROLLING

# Trolling, as tricksterism



*Are you familiar with fishing? Trolling is where you set your fishing lines in the water and then slowly go back and forth dragging the bait and hoping for a bite. Trolling on the Net is the same concept - someone baits a post and then waits for the bite on the line and then enjoys the ensuing fight.*

- Anonymous usenet poster, 1995 (relayed by Judith Donath, *Identity and Deception in the Virtual Community,* 1998)

# Trolling, as governance




*The well-constructed troll is a post that induces lots of newbies and flamers to make themselves look even more clueless than they already do, while subtly conveying to the more savvy and experienced that it is in fact a deliberate troll. If you don't fall for the joke, you get to be in on it.*

- Troller's FAQ by trollfaq@altairiv.demon.co.uk, 1996

# Trolling, as abuse



*Being a prick on the internet because you can. Typically unleashing one or more cynical or sarcastic remarks on an innocent by-stander, because it's the internet and, hey, you can.*

- EREALLY GUD DEFUNITION MAKUR, Urban Dictionary, 2004

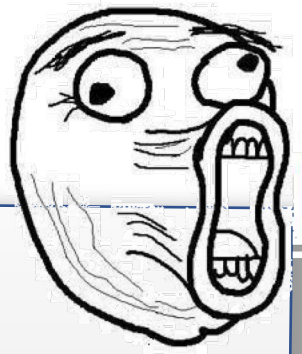
# Trolling, as security economics?



*A smart vendor treats vulnerabilities less as a software problem, and more as a PR problem. So if we, the user community, want software vendors to patch vulnerabilities, we need to make the PR problem more acute.*

- Bruce Schneier, 2007

# Trolling, as lulz




*Laughter at the expense or misfortune of others.*

- A troll to Gabriella Coleman,

*Hacker, Hoaxer, Whistleblower, Spy,* 2014

# Trolling, as transgression



Part of:

*a rich aesthetic tradition of spectacle and transgression... which includes the irreverent legacy of phreakers and the hacker underground.*

- Gabriella Coleman, *Phreaks, Hackers, and Trolls: The Politics of Transgression and Spectacle,'* 2012

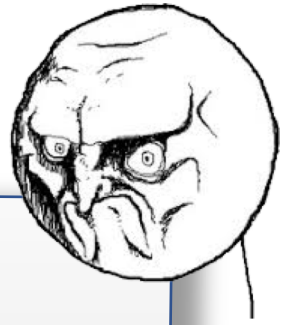
# Trolling, as art



*The best challenge to the authority of something is to find where its semantic or enforceable borders break down and to exploit those shortcomings.*

- Brad Troemel, proclaimed the “Troll of Internet Art,” 2012

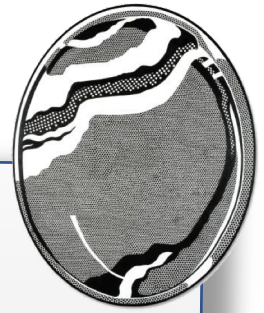
# Trolling, as hate



*We've come up with the menacing term 'troll' for someone who spreads hate and does other horrible things anonymously on the internet.*

- Adrian Chen, *MIT Technology Review,* 2014

# Trolling, as tacit media critique



*Until sensationalist, exploitative media practices are no longer rewarded with page views and ad revenue—in short, until the mainstream is willing to step in front of the funhouse mirror and consider the contours of its own distorted reflection—the most aggressive forms of trolling will always have an outlet, and an audience.*

- Whitney Phillips, *This Is Why We Can't Have Nice Things,* 2015

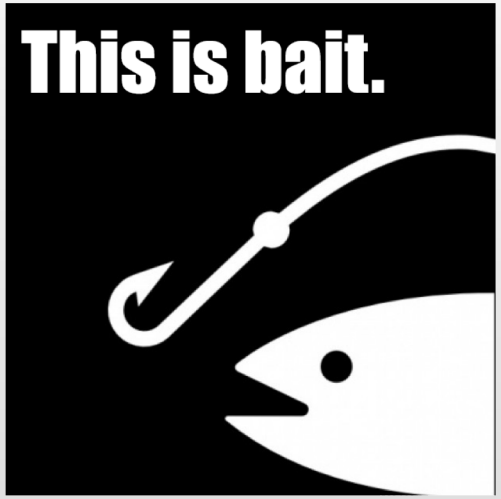
# Trolling, who?



- Users looking to govern or police their communities
- Hackers looking to corral attention or pressure vendors
- Subcultural trolls looking for ‘lulz’
- Artists looking to critique cultural practices
- Political actors looking to destabilize public sphere
- Political actors looking to corral attention or attack opponents

# Trolling, components

- Troll
- Bait
- Target
- Third-party to interpret the exchange
- Intended outcome
  - (Possible amplification)
  - (Possible political reframing and subsequent amplification)

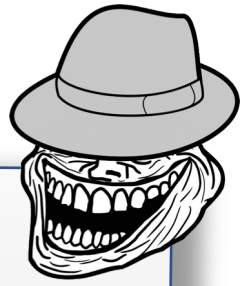


# Trolling, what now?



- “Cognitive Hacking”? (Cybenko, et al)
- “Attention Hacking”? (boyd)
- “Cognitive Denial of Service”? (Waltzman)
- Exploiting Socio-Technical “Vulnerabilities”?
- A bug or a feature?

# Trolling, how?



- Poe’s Law
- Context Collapse
- Network Anonymity and Ephemerality
- Cross Platform Ambiguity
- Attention Economics (high cost of verification)
- Lack of Trust

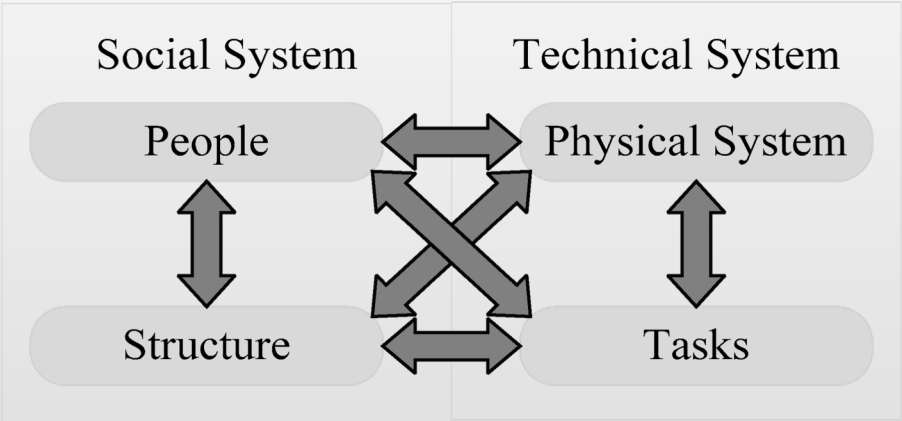
# Political Possibilities



- **Draw** attention to ignored issues (agenda setting)
- **Distract** attention from issues
- Put **pressure** on entity (security economics) (accountability)
- Draw a telling **response** (transparency)
- **Deny** a response
- **Diminish** an opponent's resources or morale

# Trolling as Socio-Technical Hacking

- Design ?vulnerabilities?
- Intent (public good) vs. Reality (profit)
- “Latent affordances”
- Trolling is to socio-technical systems what hacking is to technical systems
- What lessons are translatable?

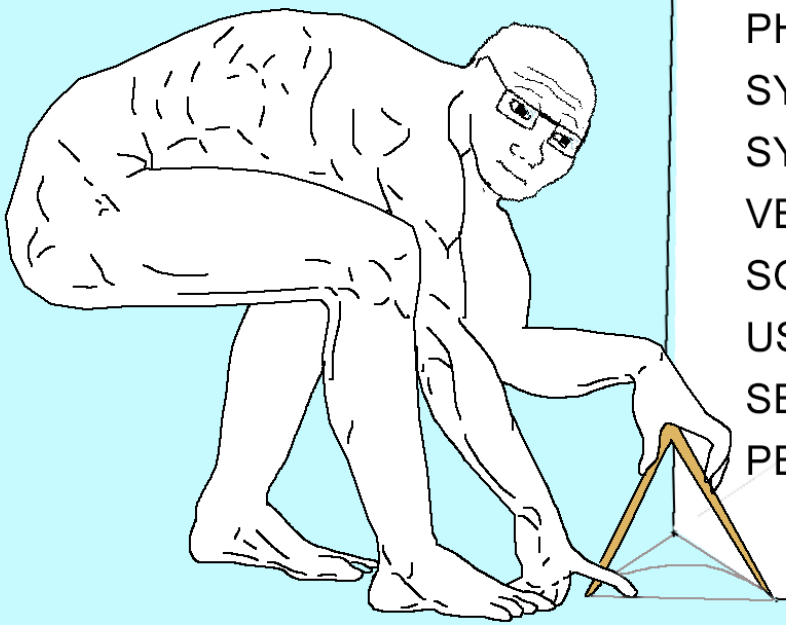


SECTION 2




# THE MEDIA AS A SECURITY ENVIRONMENT

# Maybe?



| COMPUTER SECURITY | MEMETIC SECURITY |
| --- | --- |
| SOFTWARE | MEME |
| HACKER | TROLL |
| PHISHING | BAITING |
| SYSTEM | PLATFORM / OUTLET |
| SYSADMIN | EDITOR / MODERATOR |
| VENDOR | PUBLISHER / OWNER |
| SOFTWARE DEVELOPER | MEMER |
| USER | AUDIENCE / LURKER |
| SECURITY RESEARCHER | ??? |
| PENETRATION TESTER | CRITICAL TROLL / ARTIST |
| | WHITE HAT TROLL? |

# Trolling as Attention Hacking

- Agenda Setting (from below)
- Attention is a scarce resource (attention economics)
- Legacy gatekeepers (media) compete with an emergent class of network (social media) gatekeepers to direct attention.
- Trolling as social engineering: gaining access (via gatekeepers) to the resources of a closed system (their audience)

# Network Gatekeeping

- Newspaper editors and reporters
- Platform designers / vendors
- Platform algorithms
- Content moderators
- Advertisers
- Anyone with framing or curating capacity (i.e.: Users)

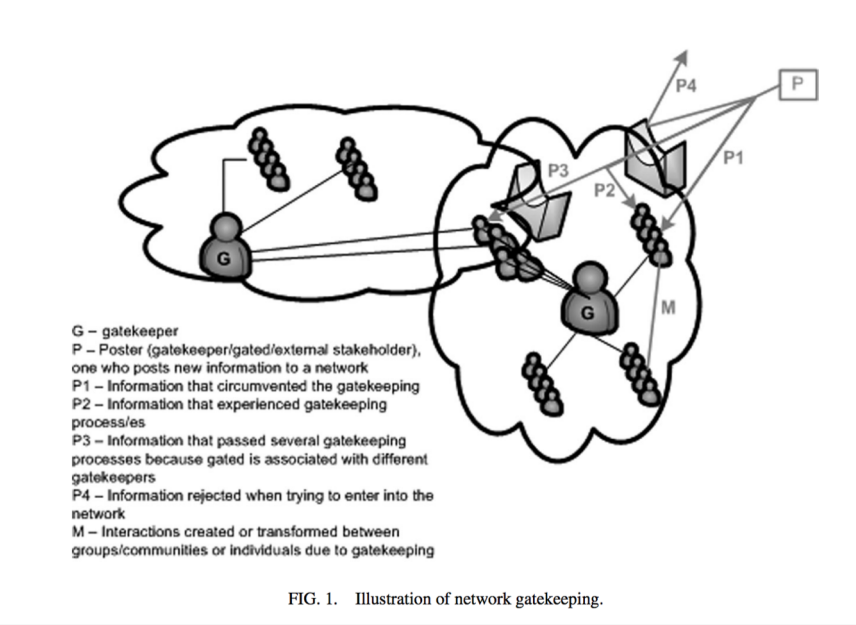



FIG. 1. Illustration of network gatekeeping.

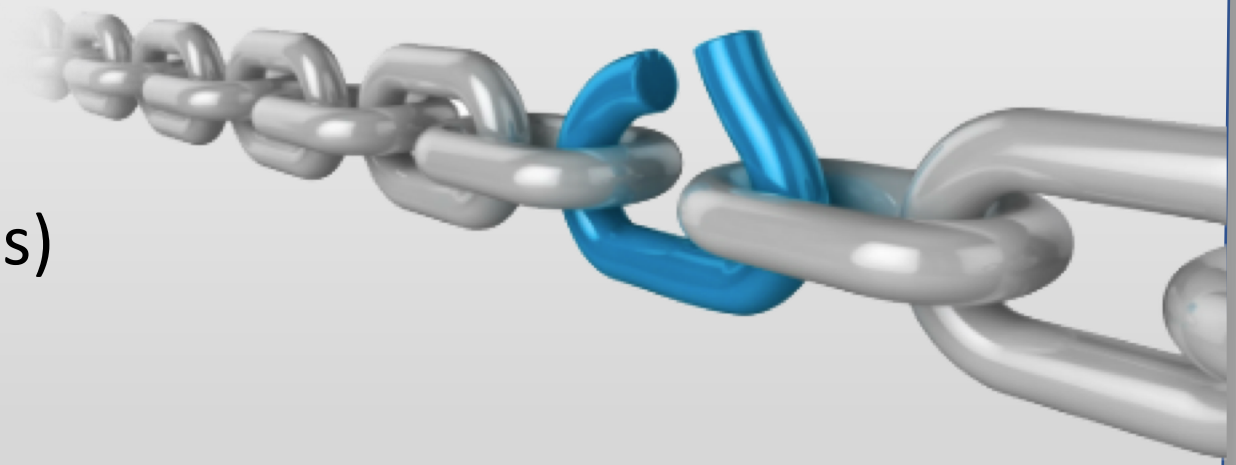
Karine Barzilai-Nahon, 2008

# Trolls

- Exploiting or detourning gatekeeper vulnerabilities to affect readers / users
- Eliciting changes to gatekeeping practices
- Disclosing dynamics inherent to gatekeeping
- Posing as gatekeepers
- Capturing legacy gatekeeper resources / audiences

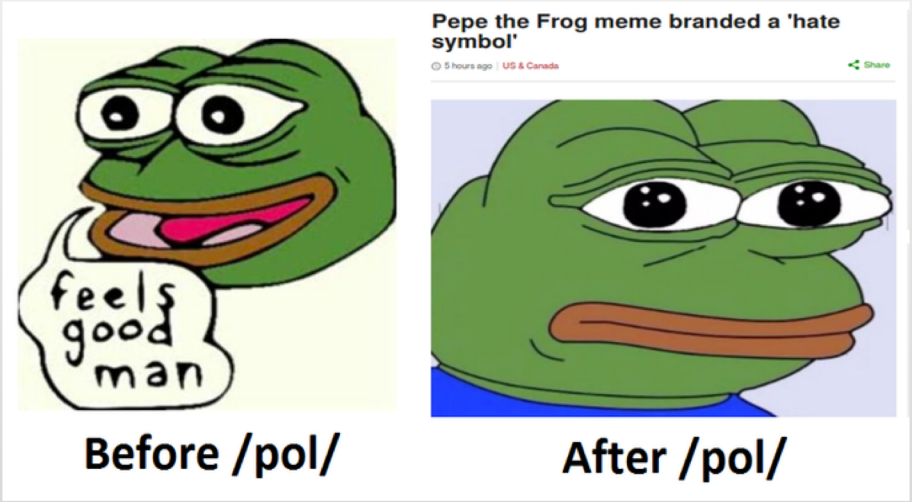
# Vulnerabilities?

- Clickbait / ”Bleeding Leads” / Breaking News
- Network curation
- Asymmetric cost to debunk
- Automated amplification (bots)
- Adtech
- Context collapse / Poe’s Law / Anonymity

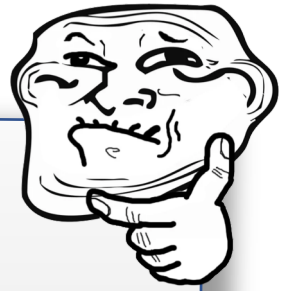


# Memes as Malware?

- Pepe as a Trojan Horse
- Vulnerabilities all the way down?
- Where does securitization stop?



# Questions



- What is gatekeeper responsibility to users?
- Is this codified? How?
- Is paternalistic gatekeeping inherently authoritarian?
- Is network gatekeeping inherently democratic?
- When is gatekeeping in the public interest?

SECTION 3







# POSTURES

# Black hat

- Non-prescribed engagement
- Push private agenda
- Counter to target’s interest



# Gray hat

- Non-prescribed engagement
- Demonstrate attack surface (proof of concept)
- Trolling as revelatory moment
- Increase pressure on gatekeepers
- Troll back
- Sense of public interest?



# White hat

- Align system with intent
- Disclose vulnerabilities selectively to platforms / journalists
- Disable possibility of non-prescribed engagement through research and education
- Reduce attack surface
- Create a discursive “patch” (i.e. Godwin’s Law)



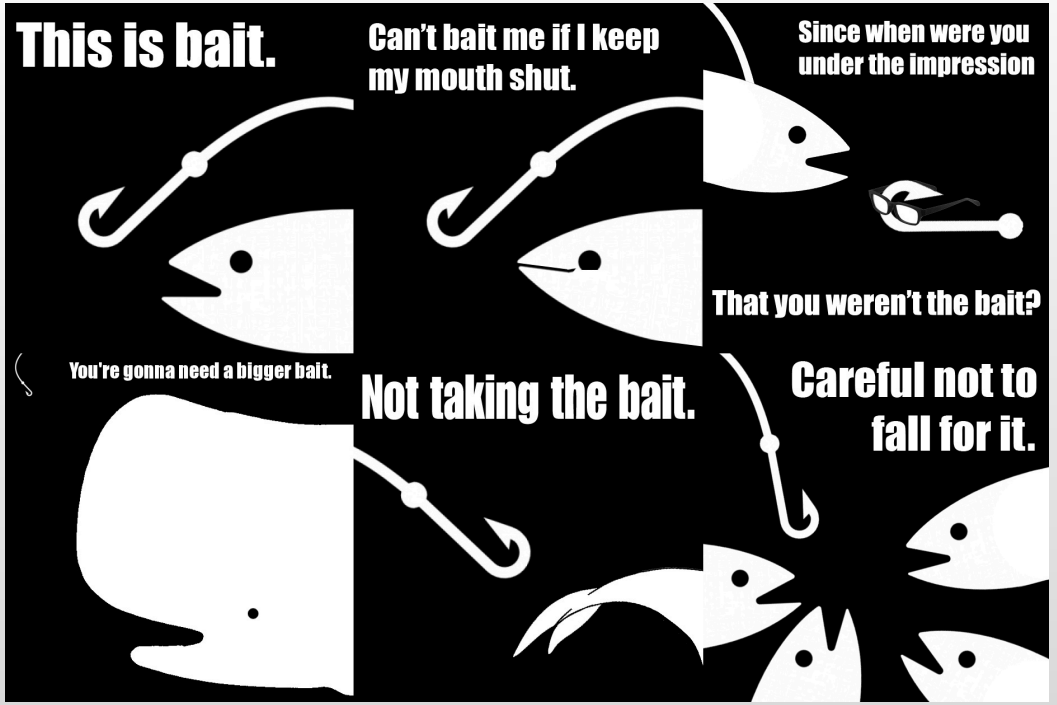
# Mechanics

- Who?
- What are the incentives?
- Who is being tasked to respond?
- What is the form of disclosure?
- Whose interest is served?

# And the lulz?

- How do ”the lulz” figure in?
- What are the ethics of fun at others’ expense?
- Has the adoption of trolling as a propaganda technique by states and political radicals changed the possibility of just “doing it for the lulz”?  Hacking for fun?

## SECTION 4




# INVENTORIES

# Possible Goals



- Political Maneuvering
- Chaos / Scorched Earth
- Perception Management
- Denial of Service
- Lulz / Drama
- Agenda Setting / Priming
- Incitement
- Response seeking
- Community building
- Community governance
- Demoralization / Outrage
- Profit???

# Tactics



- Source Hacking / Journobaiting
- Attentional Honeypots
- Keyword Squatting
- Capturing the Narrative
- Cognitive Denial of Service
- Brute Force Harassment
- Ironical Bait & Switch
- Gaming Databases / Algorithmic Control
- Memetic Trojan Horses
- Controlled Opposition
- Driving a Wedge
- Concern Trolling
- Brigading / Dogpiling
- Sockpuppetry
- Jiu Jitsu / Self-Victimization
- Doxing
- Context Distortion
- Deep Fakes / Photoshops

SECTION 5

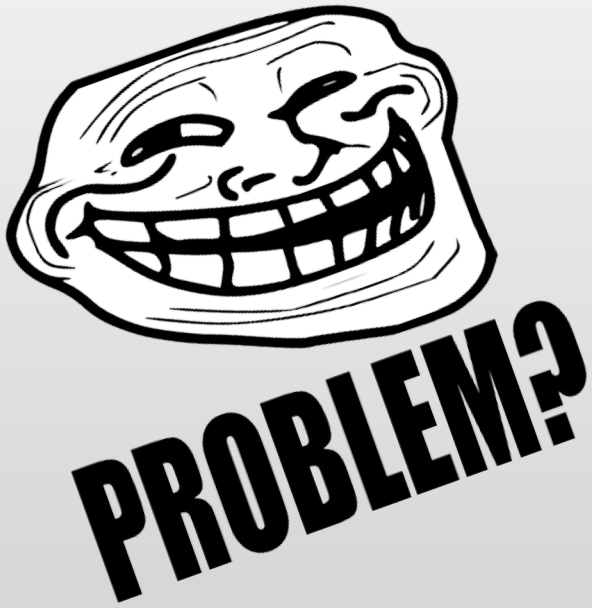


# QUESTIONING

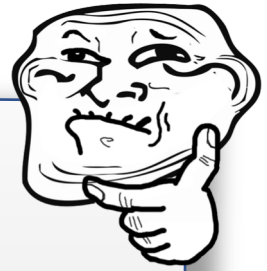
# BIG QUESTION:

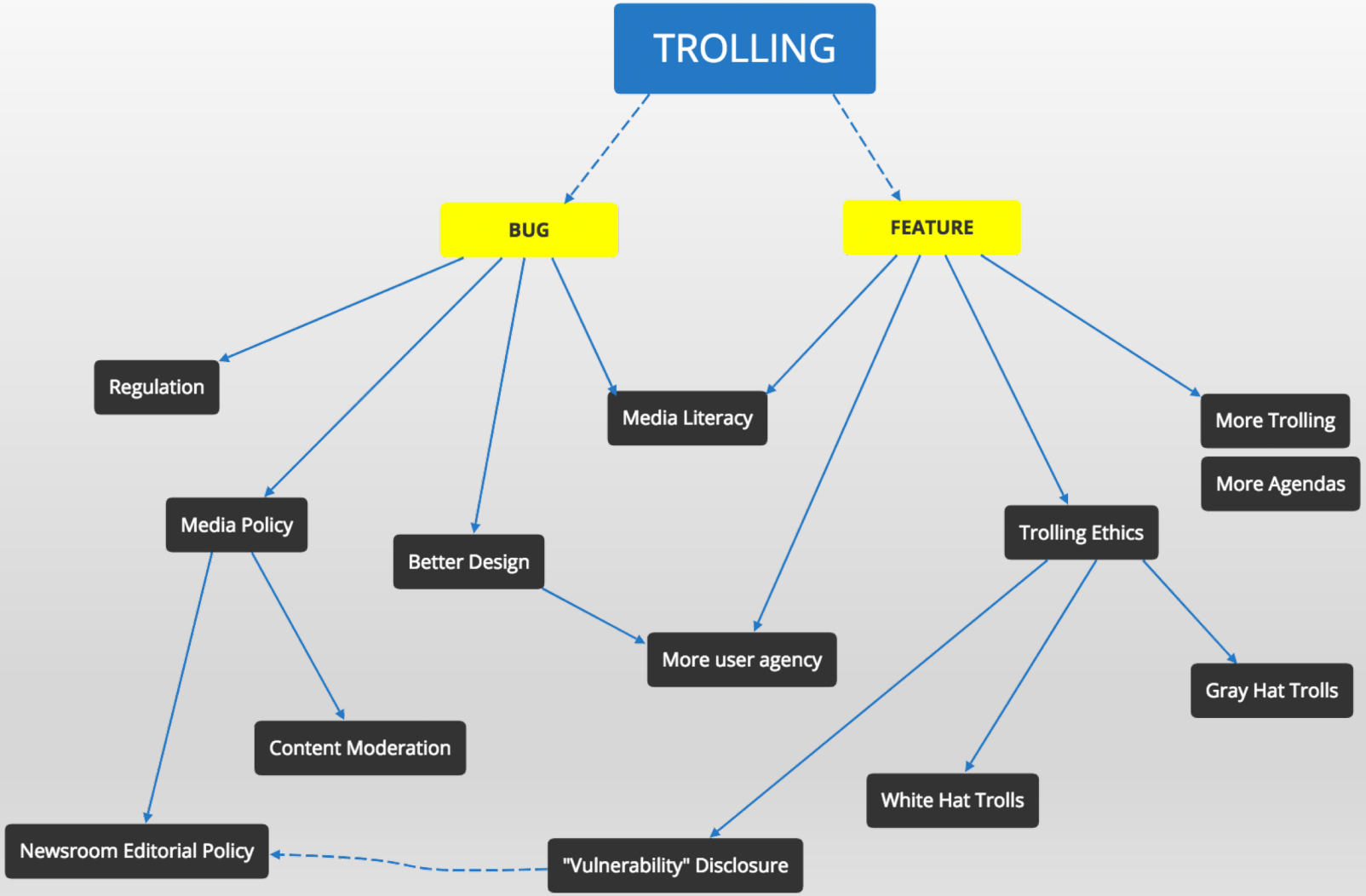
- **A bug or a feature?**
- Is trolling a problem?
- Or is this how networked media dynamics are supposed to work?

**and WHO DECIDES?**



# BUG vs. FEATURE:





# “Feature” Interventions:



- Do nothing
- Incentivize trolling from diverse perspectives
- A trolling arms race
- A trolling code of ethics

**BUT?**

# “Bug” Interventions:



- Technical re-design
- Policy change
- Content moderation
- Media literacy (for users *and* gatekeepers?)
- Norms like “Don’t Feed the Trolls” / “Strategic silence”
- Regulation
- Institutionalized trust and verification

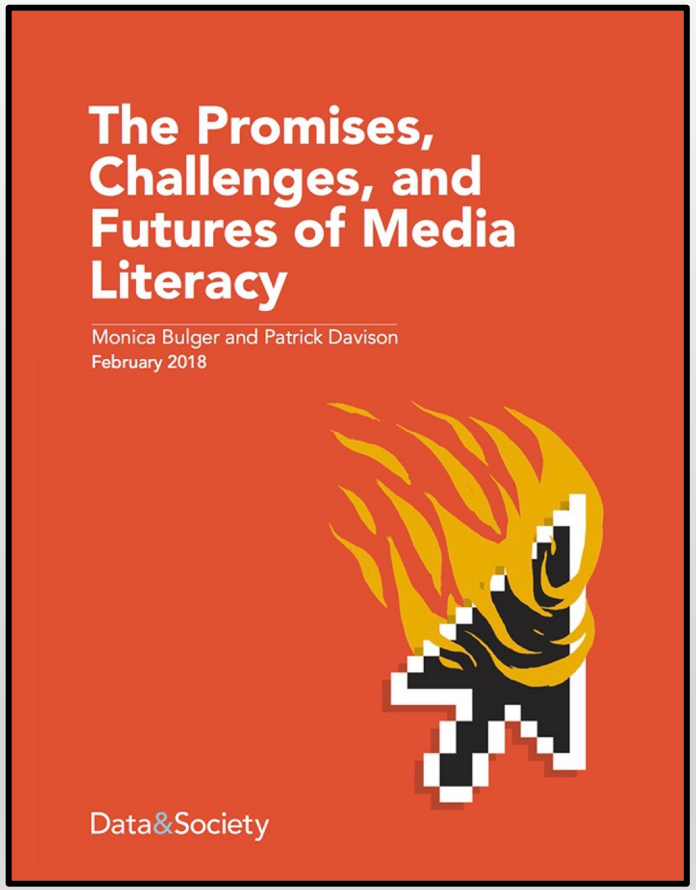
**BUT?**

# To Consider:

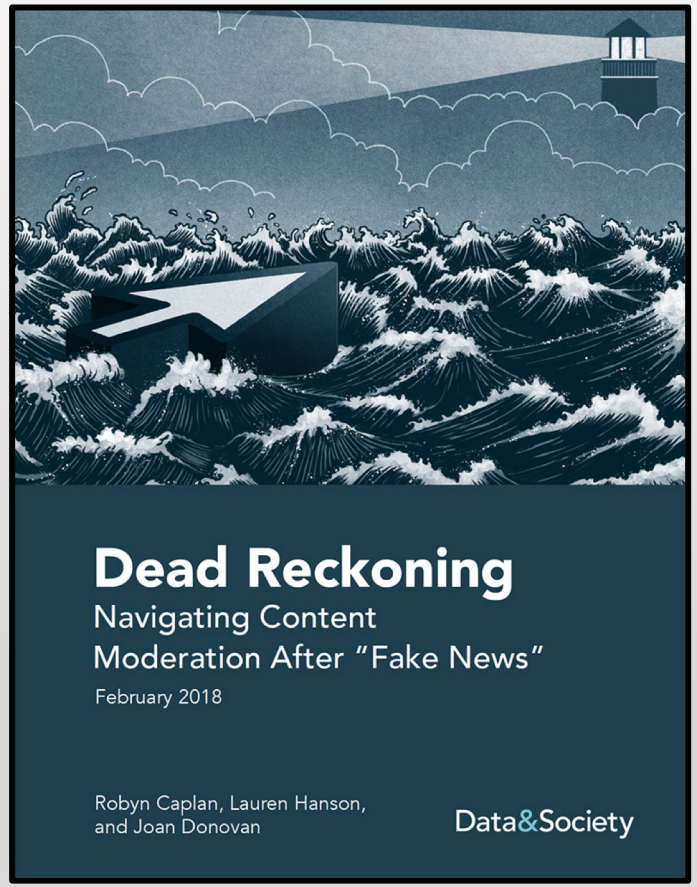


- Right to speech ≠ Right to amplification?
- Media Literacy vs. Attention Economics?
- Network Gatekeeping vs. Paternalism?
- What Public Interest? What Ethical system? Which Establishment?
- Who discloses, how, to whom, to what end?

# Interventions

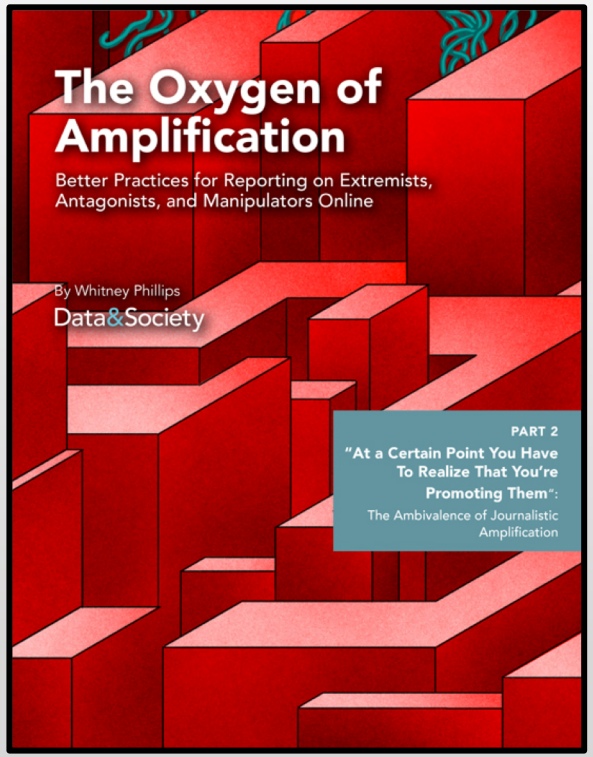


# Interventions



# Interventions






# Free Speech?



- Right to free speech with your one mouth?
- Right to amplification of that speech?
- Monopolization of attention = DOS of other's free speech?

# Everyone an investigative journalist?



- Finite lifetime problem
- Can’t verify/investigate every claim
- Inflammatory lies stickier and more profitable

# Moderation?



- Moderation under counts
- No moderation over counts
- Do-it-yourself moderation tends to tribalism?

# Regulation?



- Lessons from Computer Fraud and Abuse Act?
- Chilling Effect?
- Trolls, Platforms, Journalists?
- Negative externalities?
- Current laws in place?

# Platform Redesign?



- To who's standards?
- Addressing what issues?
- Informed by red teaming "white hat trolls?"
- Informed by external pressure from "gray hat trolls?"

# Disclosure?

- Proof of Concept?
- Full disclosure? Security Economics?
- Markets?
- Selective Disclosure? Security Research?
- Bounty Programs? Penetration Testing? Auditing?
- Professionalization?





# More Trolls?

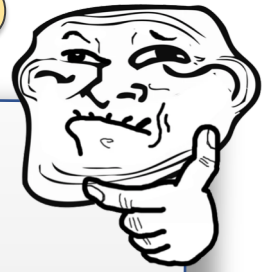
- White Hats, Gray Hats, Black Hats?
- More Positions, More Issues
- Politics of Spectacle
- Agonistic Politics
- Race to the Bottom?
- Norms or Codes of Ethics?





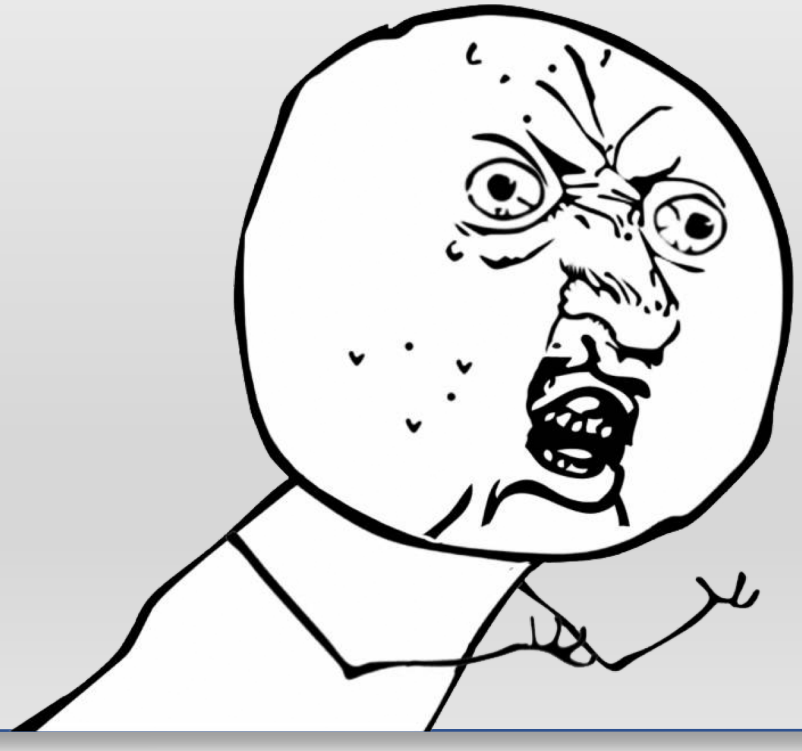


# Which brand of wrong do we want?



All of the above but none too well – cracks are where the light gets through!

# DISCUSSION



# CONCLUSIONS

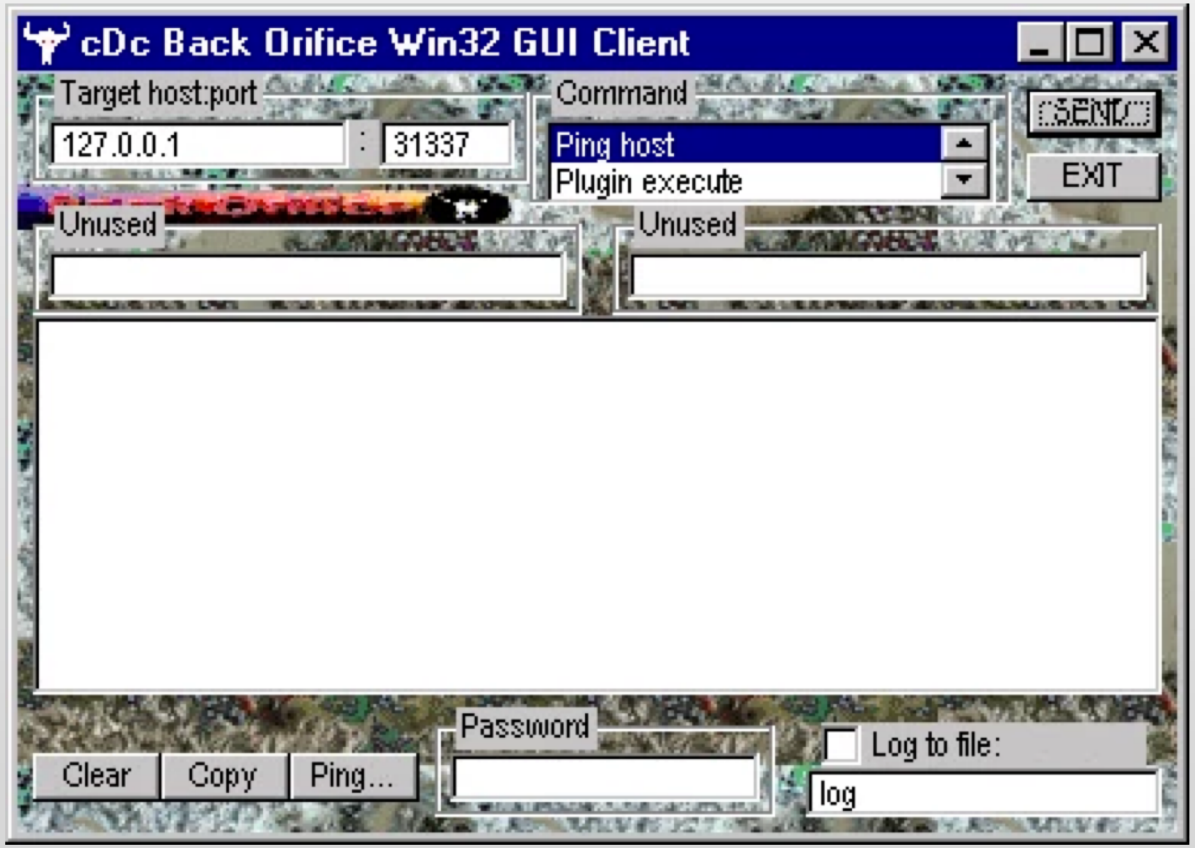


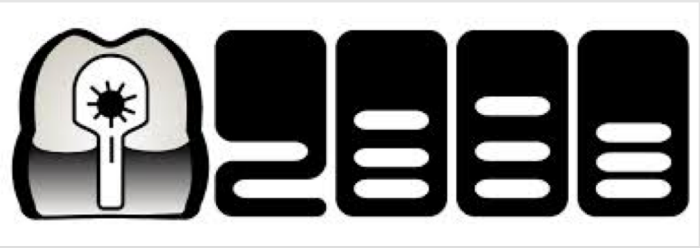
## APPENDIX



# CASE STUDIES

# BACK ORIFICE



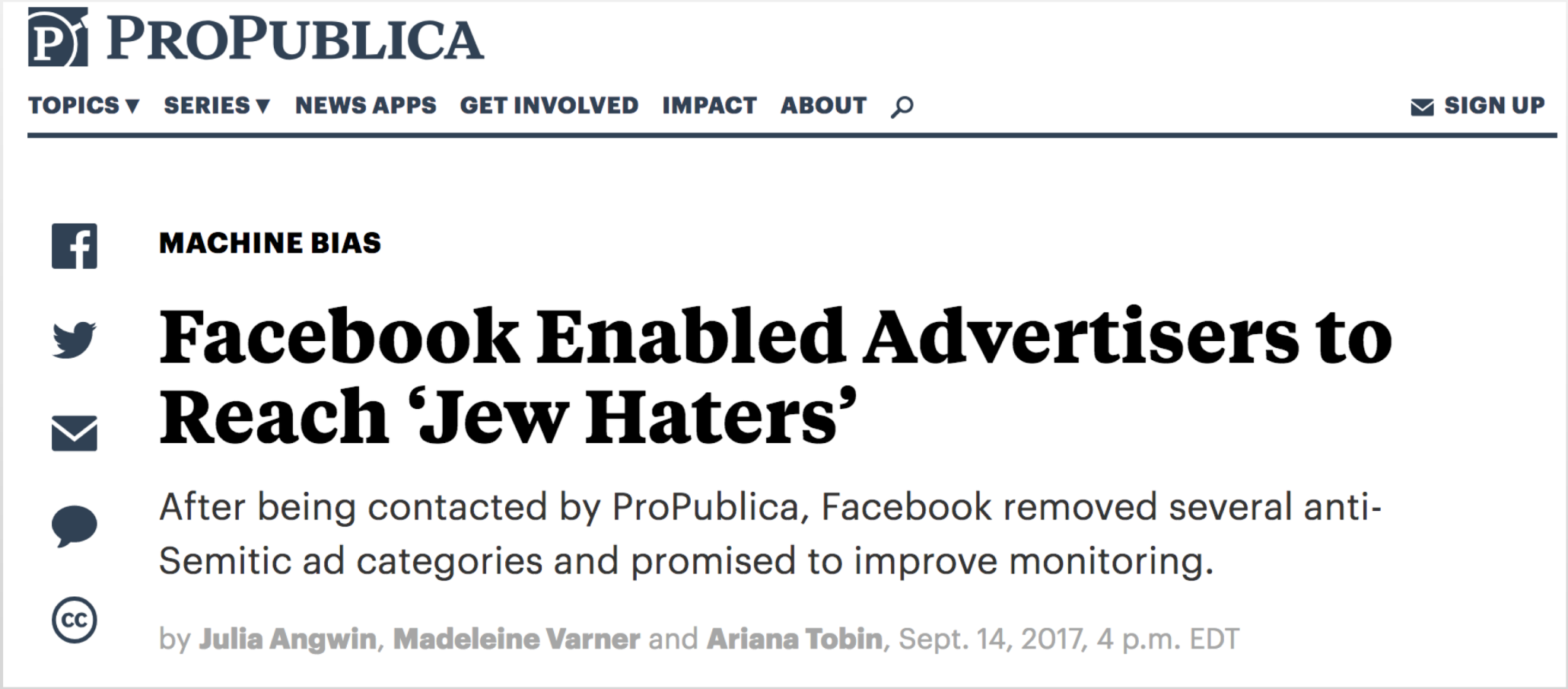
- Security Economics

# HACKED TEAM




- Politics of Spectacle / Security Economics

# PROPUBLICA INVESTIGATION



PROPUBLICA

TOPICS ▾ SERIES ▾ NEWS APPS GET INVOLVED IMPACT ABOUT SIGN UP

MACHINE BIAS

## Facebook Enabled Advertisers to Reach 'Jew Haters'

After being contacted by ProPublica, Facebook removed several anti-Semitic ad categories and promised to improve monitoring.

by **Julia Angwin**, **Madeleine Varner** and **Ariana Tobin**, Sept. 14, 2017, 4 p.m. EDT

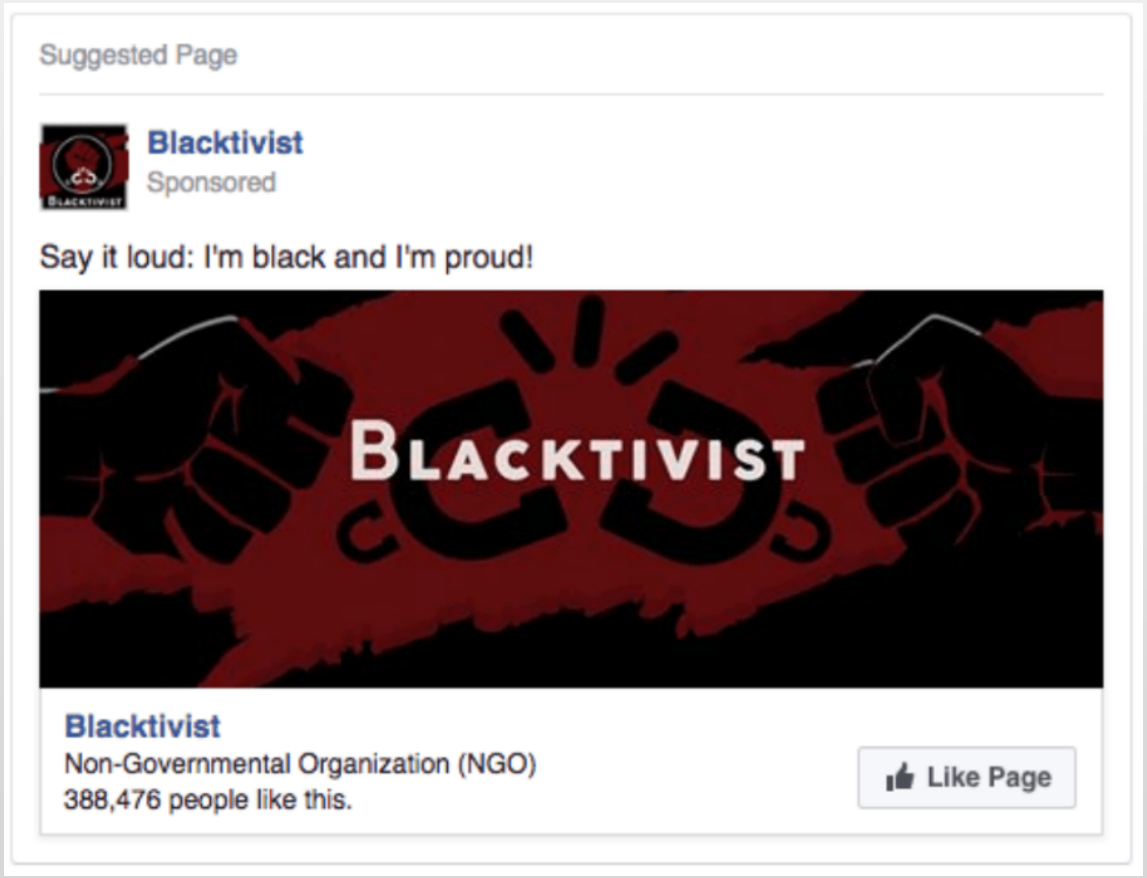
- Security Economics

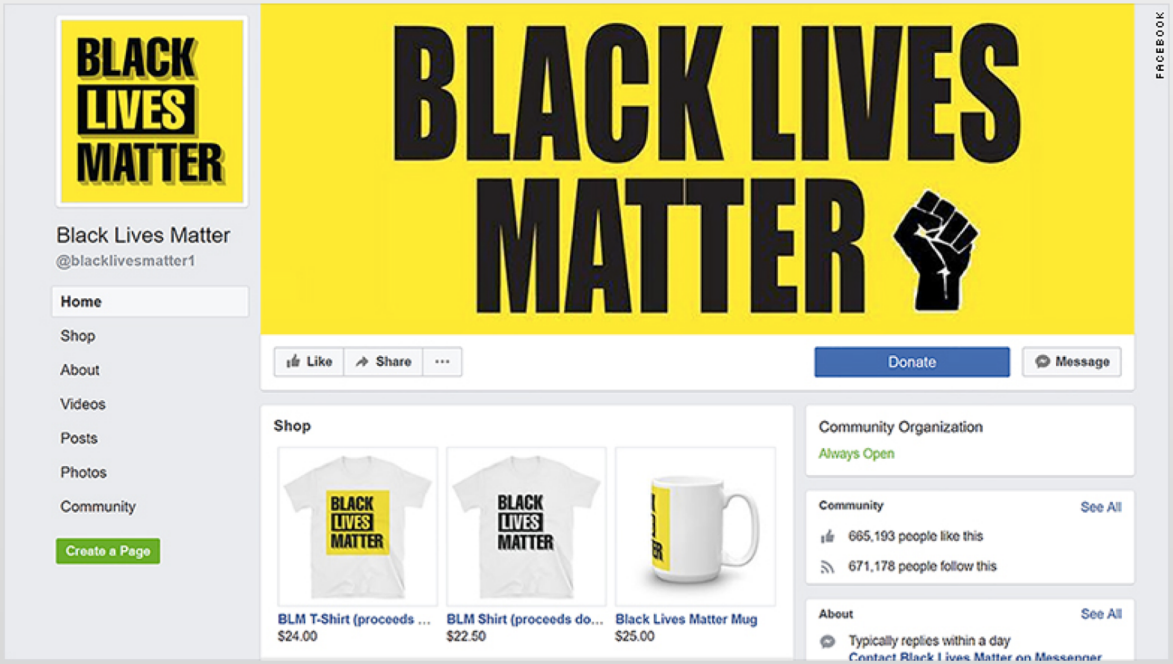
# ANTIFA SOCKPUPPETS




- Keyword Squatting

# “BLACKTIVIST”




- Keyword Squatting

# INTERNET RESEARCH AGENCY




- Scorched Earth

# TAY AI




- Algorithmic Gaming

# DOXING NSA GRAPHIC DESIGNER




- Critical Trolling

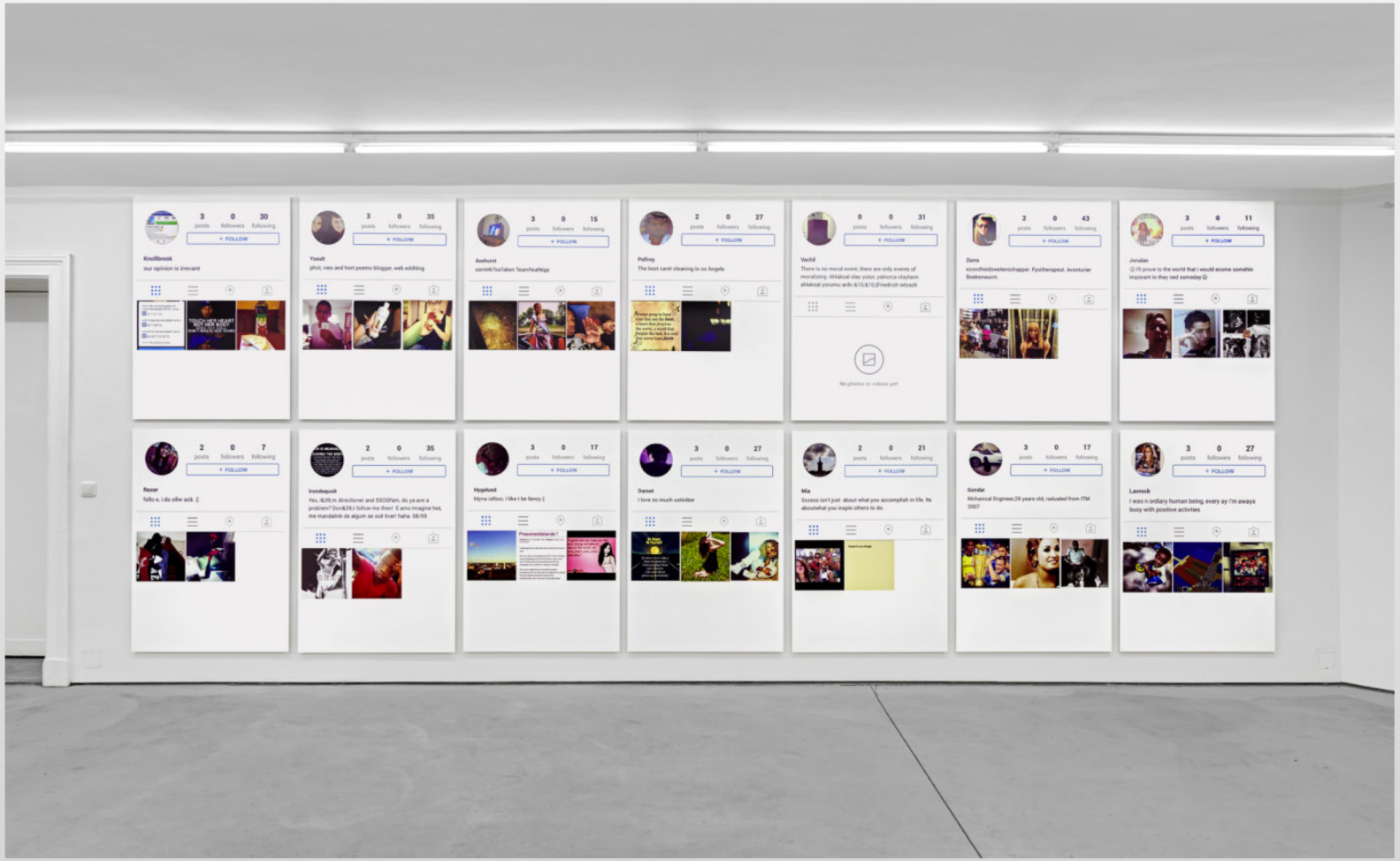
# TITS.AI and ROCKET AI




- Critical Trolling

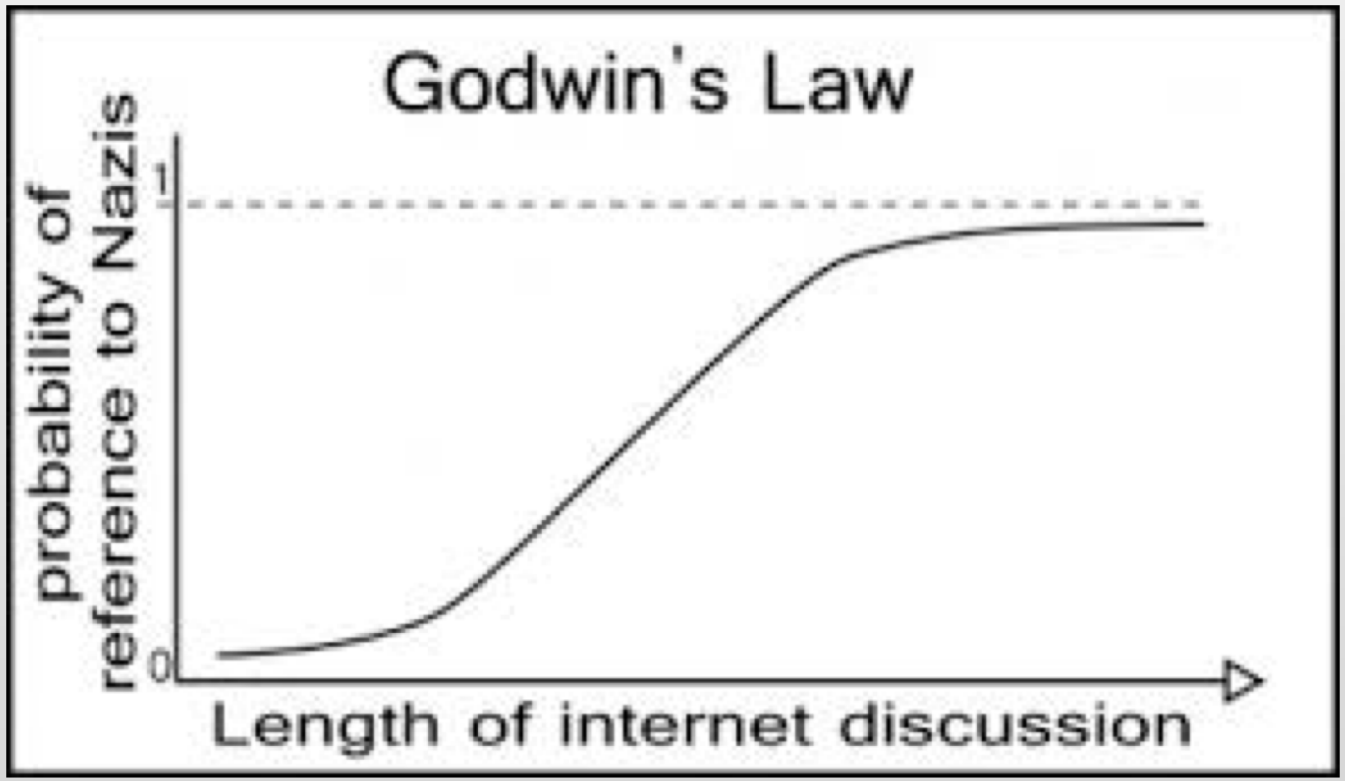
# 100,000 FOLLOWERS



- Critical Trolling

# GODWIN'S LAW



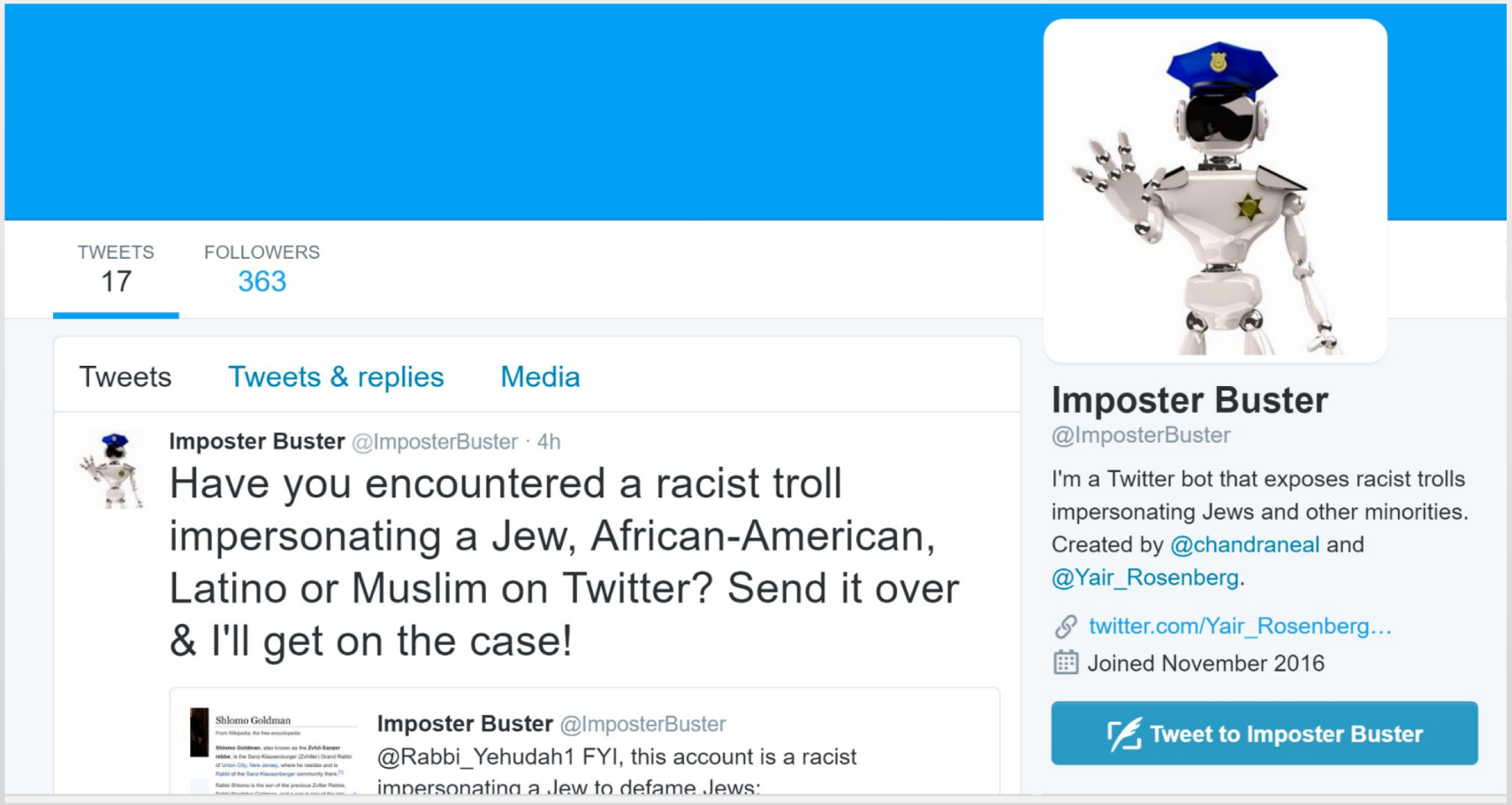

- Counter Meme

# SAM HYDE IS THE SHOOTER




- Source Hacking

# IMPOSTER BUSTER




- Counter Trolling

# OVER 9000 PENISES




- Lulz or Retroactive Proof of Concept

# POOL'S CLOSED




- Lulz or Proof of Concept

# PROMOTED TWEETS




- Attention Hacking

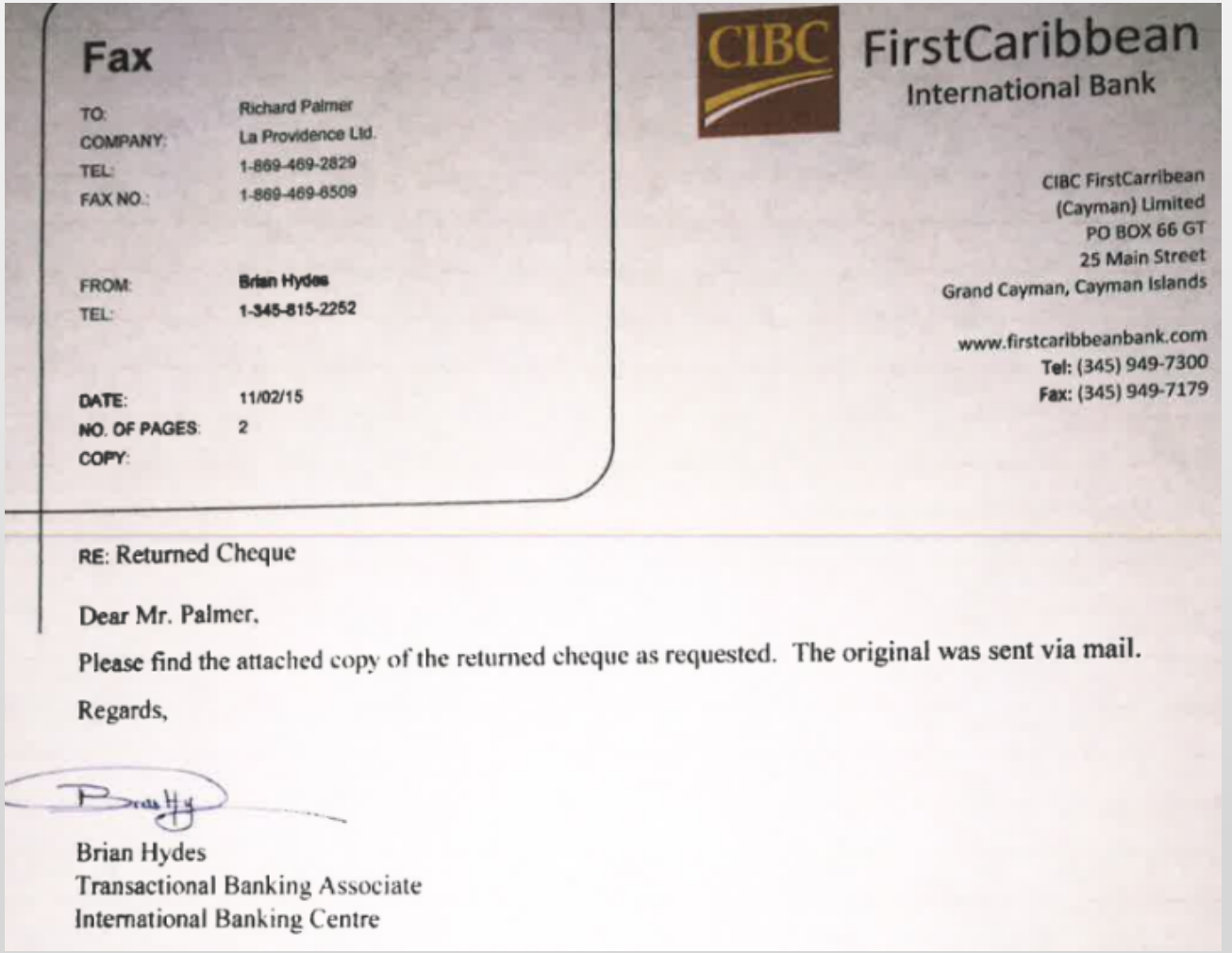
# SHAQ ATTACK

Are you Shaquille O'Neal - Want to see your dream artwork?
www.bocagallery.biz/shaq
Shaq saves 20%!



- Critical Trolling

# MACRON "LEAKS"



**Fax**

TO: Richard Palmer
COMPANY: La Providence Ltd.
TEL: 1-869-469-2829
FAX NO.: 1-869-469-6509

FROM: Brian Hydes
TEL: 1-345-815-2252

DATE: 11/02/15
NO. OF PAGES: 2
COPY:

CIBC FirstCaribbean International Bank

CIBC FirstCarribean
(Cayman) Limited
PO BOX 66 GT
25 Main Street
Grand Cayman, Cayman Islands

www.firstcaribbeanbank.com
Tel: (345) 949-7300
Fax: (345) 949-7179

RE: Returned Cheque

Dear Mr. Palmer,

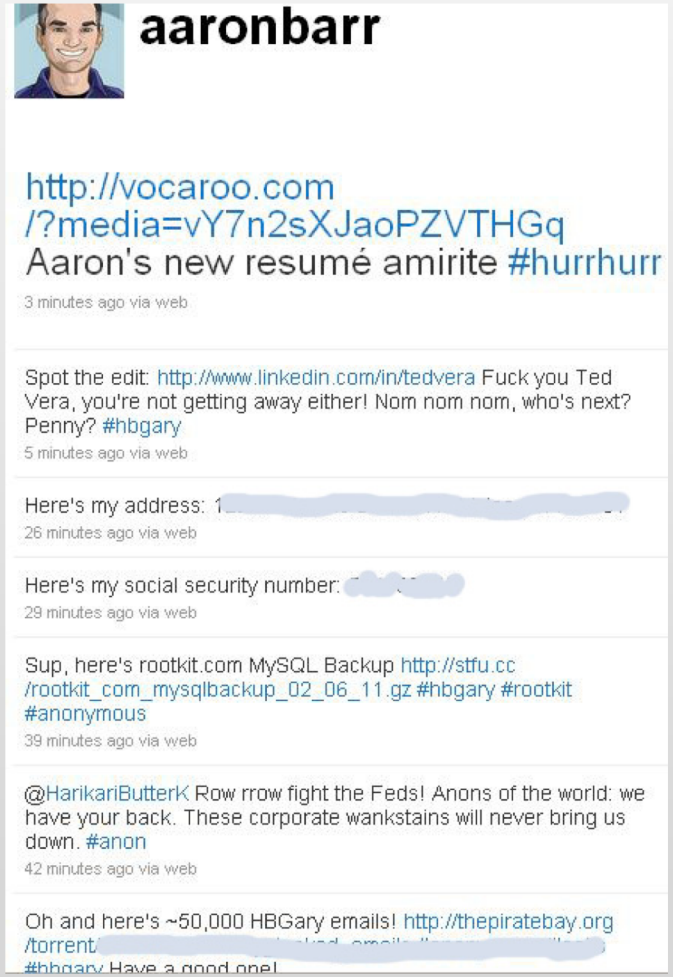
Please find the attached copy of the returned cheque as requested. The original was sent via mail.

Regards,

Brian Hydes
Transactional Banking Associate
International Banking Centre

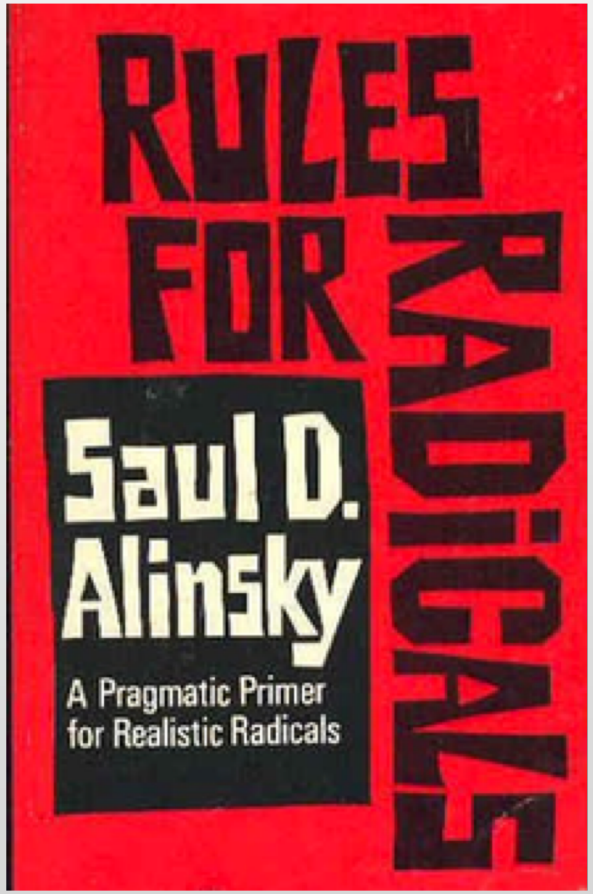
- Cognitive Denial of Service

# HBGARY




- Ruin Life Campaign

# SAUL ALINSKY



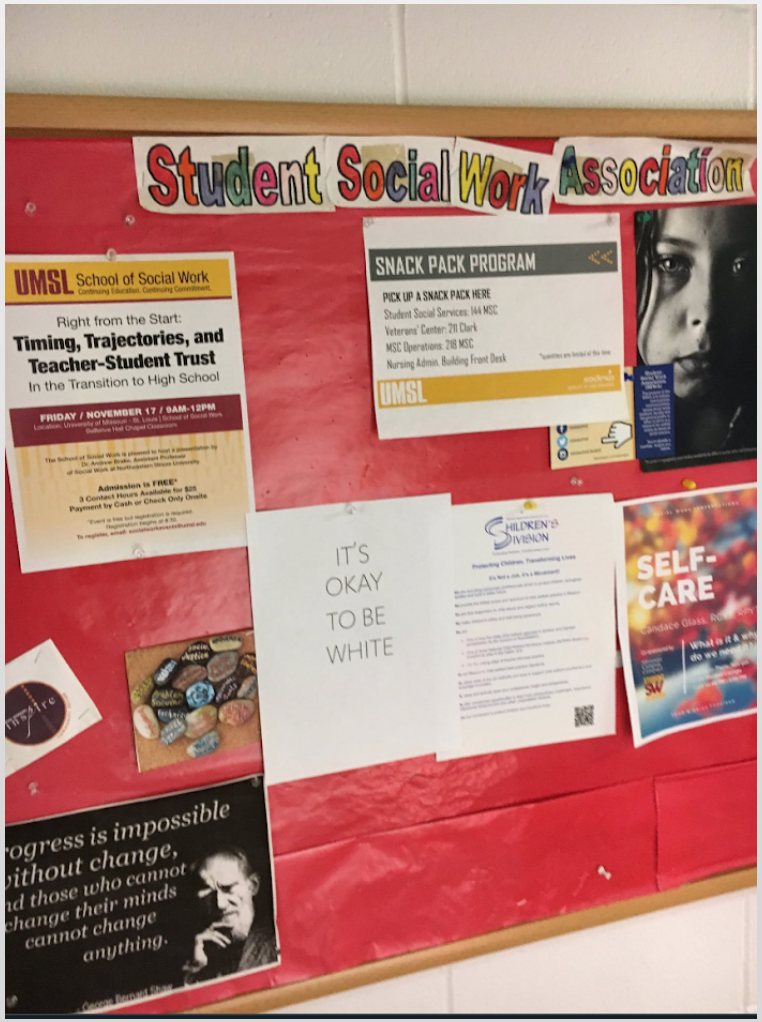

- Attention Hacking

# THE DAILY STORMER




- Weaponized Irony

# IOTBW POSTER CAMPAIGN







- Jiu Jitsu

# REPUBLIC OF FLORIDA



158KiB, 1542x1134, Me and nick.jpg

View Same | Google | iqdb | SauceNAO

**Nikolas Cruz shooting Anonymous** ID:NIYXoxqv Wed 14 Feb 2018 23:24:58 No.160524925 [126 / 35]

View | Reply | Original | Report

Quoted By: >>160525060 >>160525081 >>160525157 >>160525215 >>160525336 >>160525618 >>160525701 >>160525771 >>160526253 >>160526423 >>160526877 >>160526928 >>160528519 >>160528550 >>160528610 >>160529481 >>160529822 >>160530163 >>160530342 >>160530440 >>160530960 >>160531037 >>160531156 >>160531417 >>160532208 >>160532661 >>160533641

I've kept quiet for long enough. I've decided that its time the world know the truth about ROF.

Nikolas Cruz was a revolutionary member of the Republic of Florida, who preached twisted and dark things like terrorism and attacking innocent people.

I used to be a member of ROF, but I left because of the violent views that Nikolas and people like him were pushing. I voiced my concerns to some of the leadership of ROF, but they all supported what he said.

The Republic of Florida is terrifying and clearly they're willing to kill anyone. Nikolas was actually one of the less violent ones.

This is going to happen again if something doesn't change

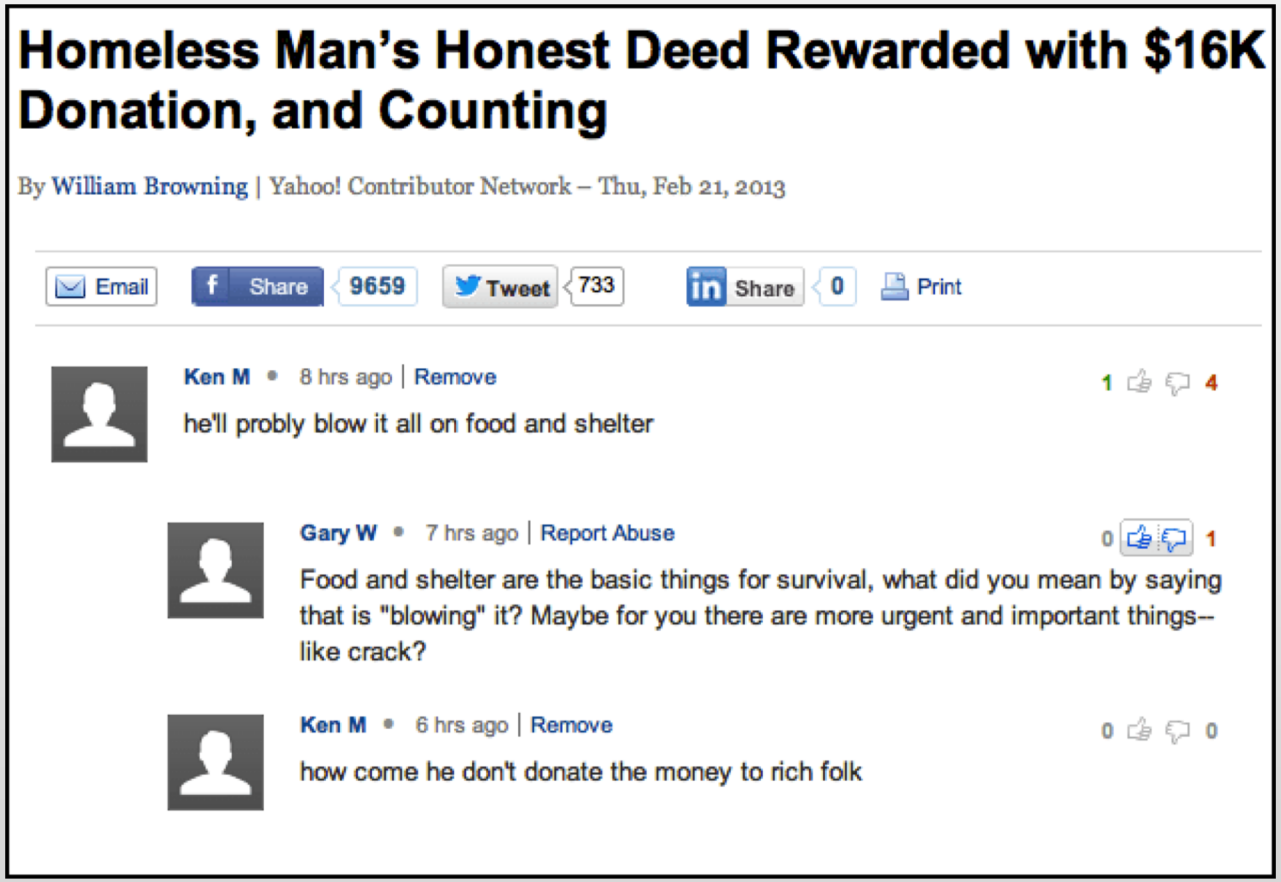
- Source Hacking

# SPOON SCULPTURE AT PURDUE PHARMA




- Attention Hacking

# KEN M – YAHOO! TROLL



**Homeless Man's Honest Deed Rewarded with $16K Donation, and Counting**

By William Browning | Yahoo! Contributor Network – Thu, Feb 21, 2013

Email | Share 9659 | Tweet 733 | Share 0 | Print

**Ken M** • 8 hrs ago | Remove — 1 4

he'll probly blow it all on food and shelter

**Gary W** • 7 hrs ago | Report Abuse — 0 1

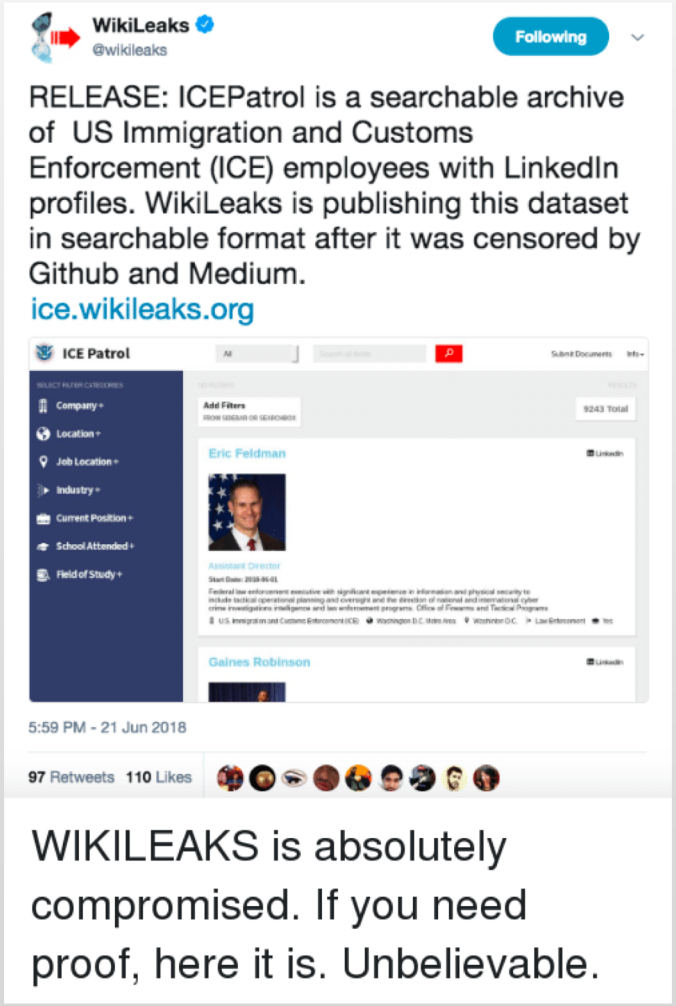
Food and shelter are the basic things for survival, what did you mean by saying that is "blowing" it? Maybe for you there are more urgent and important things--like crack?

**Ken M** • 6 hrs ago | Remove — 0 0

how come he don't donate the money to rich folk

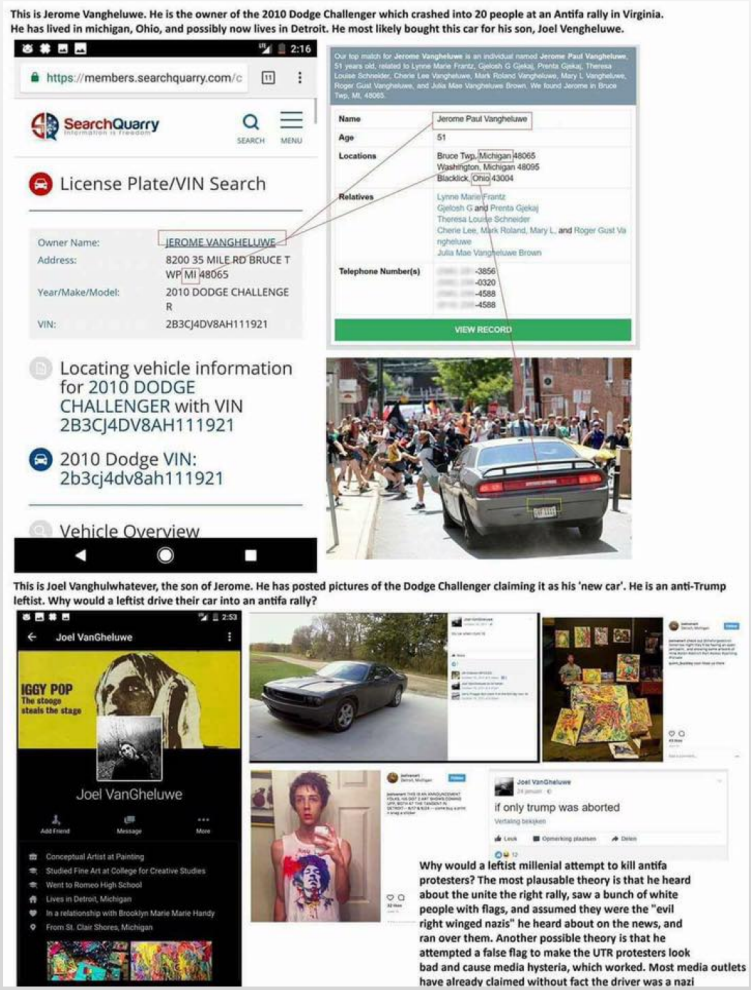
- Lulz?

# ICE DATABASE




- Brute Force Provocation

# CHARLOTTESVILLE DRIVER




- News Spamming / Counter Framing

# @YesYoureRacist




- Concern Trolling

# VIVIAN JAMES



- Concern Trolling

# OCCIDENTAL COLLEGE FAKE RAPE REPORTS



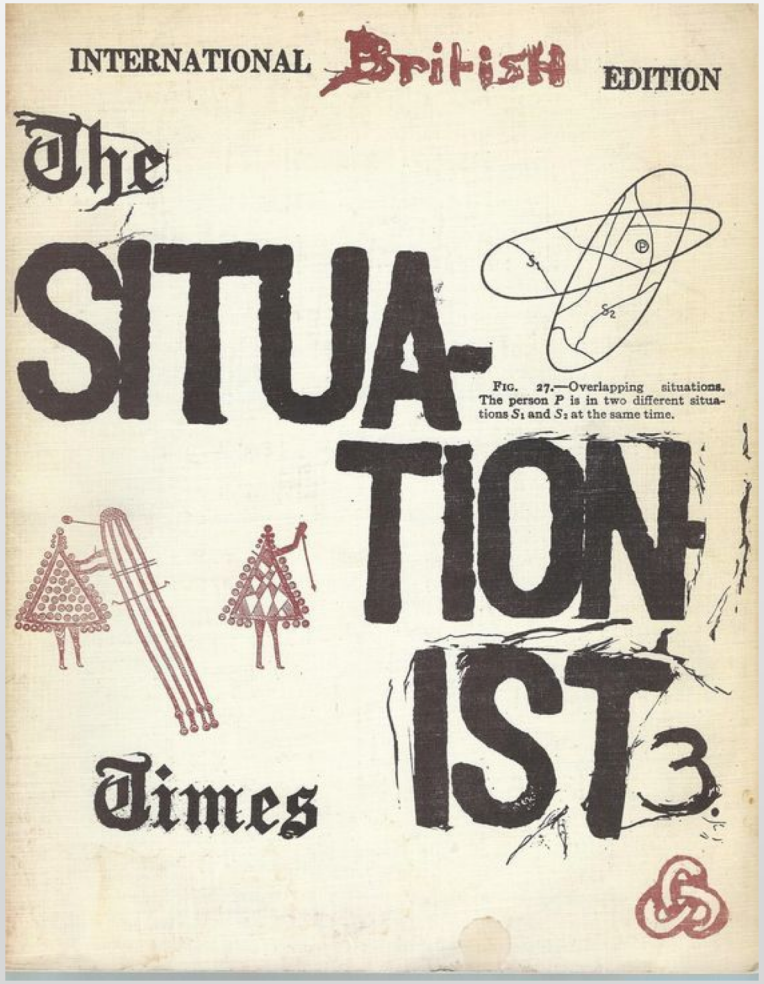
- Denial of Service / Harassment

# ESTEBAN WINSMORE



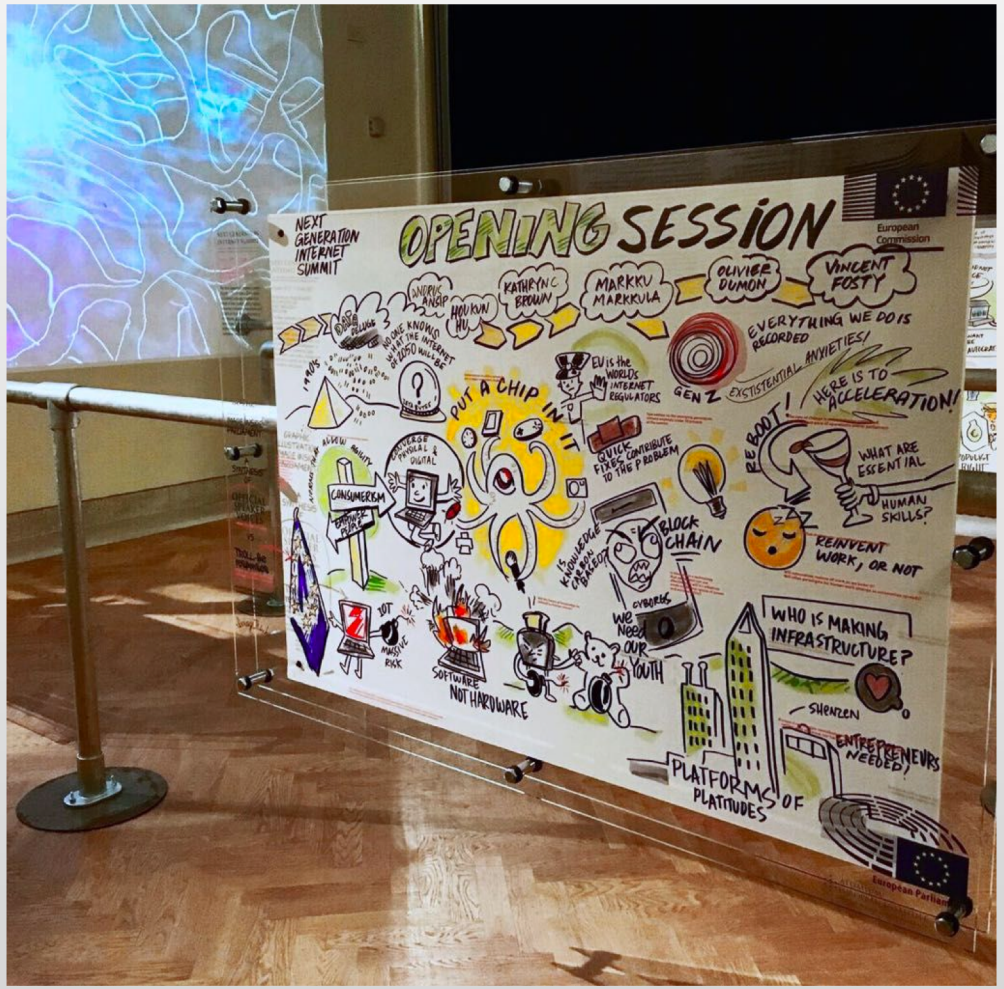

- Concern Trolling / Ethnography

# LETTRISM, SITUATIONISM




- Politics of Spectacle

# EUROPEAN COMMISSION




- Approved Trolling

# RUSSIAN MEMES

48. During November 2016, the Conspirators used Guccifer 2.0 to "shit-post" over 16,000 times online. The posts contained both domestic and foreign memes, as well as 514 separate incidences of a post consisting solely of "lol" or "lmao."

a. On or about November 18, 2016, the Conspirators created a meme known as "Borgar," which was then combined with existing viral media to infiltrate the international meme community.

b. The Conspirators did, on or about November 28, 2016, create the meme "Soup time," in which a frog steals money from his mother, diverting the funds to instead purchase

**drewtoothpaste**
@drewtoothpaste

What the FUCK to the Russian-meme details of this FBI indictment

12:40 PM - Jul 13, 2018

♡ 5,366 💬 1,649 people are talking about this

**mieke eoyang** @MiekeEoyang · 13 Jul

So, @drewtoothpaste, you want to own up to

a) a mistake
b) a joke
c) being complicit in a disinformation campaign?
twitter.com/gabrielmalor/s…

**drewtoothpaste**
@drewtoothpaste

i'm a humor writer, not a news source. don't get mad at me if you don't know who i am or don't bother to read my bio

1:59 PM - Jul 13, 2018

♡ 985 💬 61 people are talking about this

- Humor as Incidental Troll