

Finding Fixes for Fake News

Marshall Van Alstyne
Boston University

NEWSPAPERS



MAN
LANDS
ON
MARS

MARS
LANDS
ON
MAN

FICTION

NON-FICTION

THE VERY NEAR FUTURE



**FAKE
NEWS**

FACTS

**The
Economist**

NOVEMBER 6TH-10TH 2017

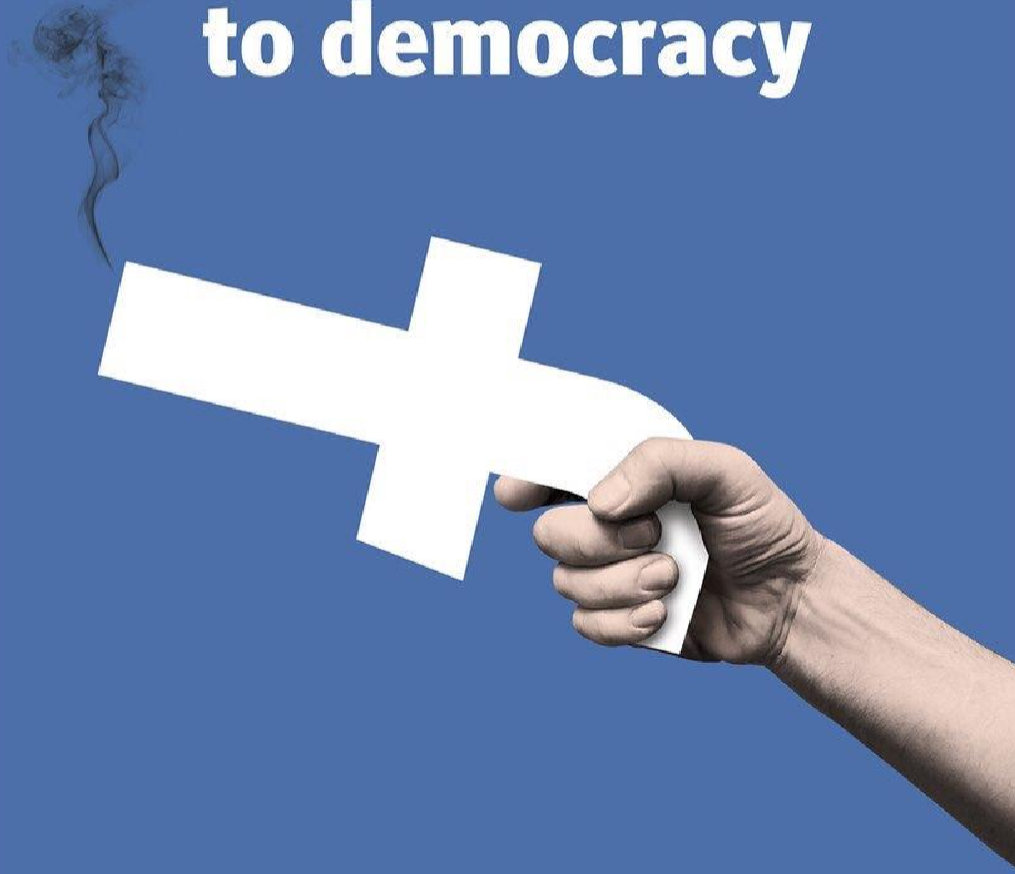
Mueller's misunderstood mission

Where next for Catalonia?

Trafficking brides to China

Luther at 500: an essay

Social media's threat to democracy



What is Fake News

? Falsehood or Misinformation



Russian troll account, TENGOP had 136,000 Twitter followers compared to actual Tennessee GOP which had 13,000

These are fans at a Cleveland Cavaliers game

What is Fake News

? Falsehood or Misinformation

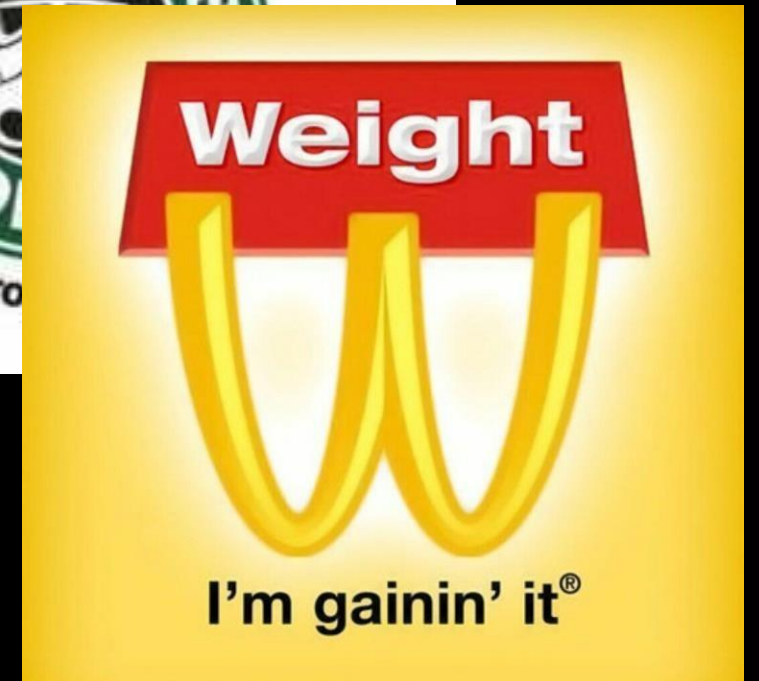
? Propaganda



Putin has been shaping public opinion at home and abroad for years. Social Media are just the new battleground

What is Fake News

- ? Falsehood or Misinformation
- ? Propaganda
- ? Parody



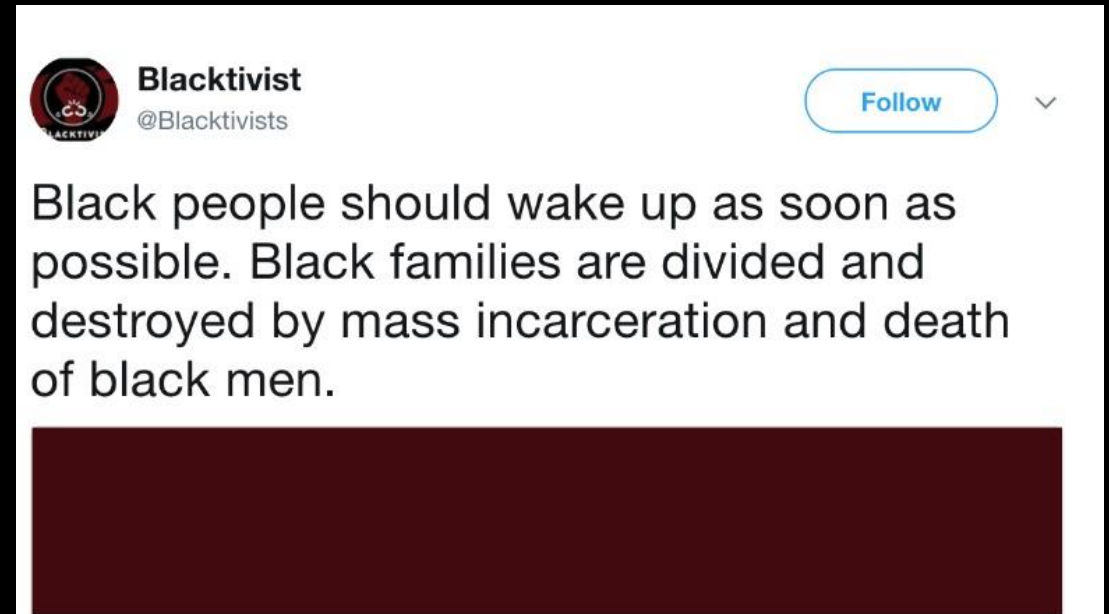
What is Fake News

? Falsehood or Misinformation

? Propaganda

? Parody

? Intentional Controversy



Russians used FB accounts to pay for black self-defense classes in tight vote areas and organize protests to inflame racial tensions. Blacktivist had 360,000 followers.

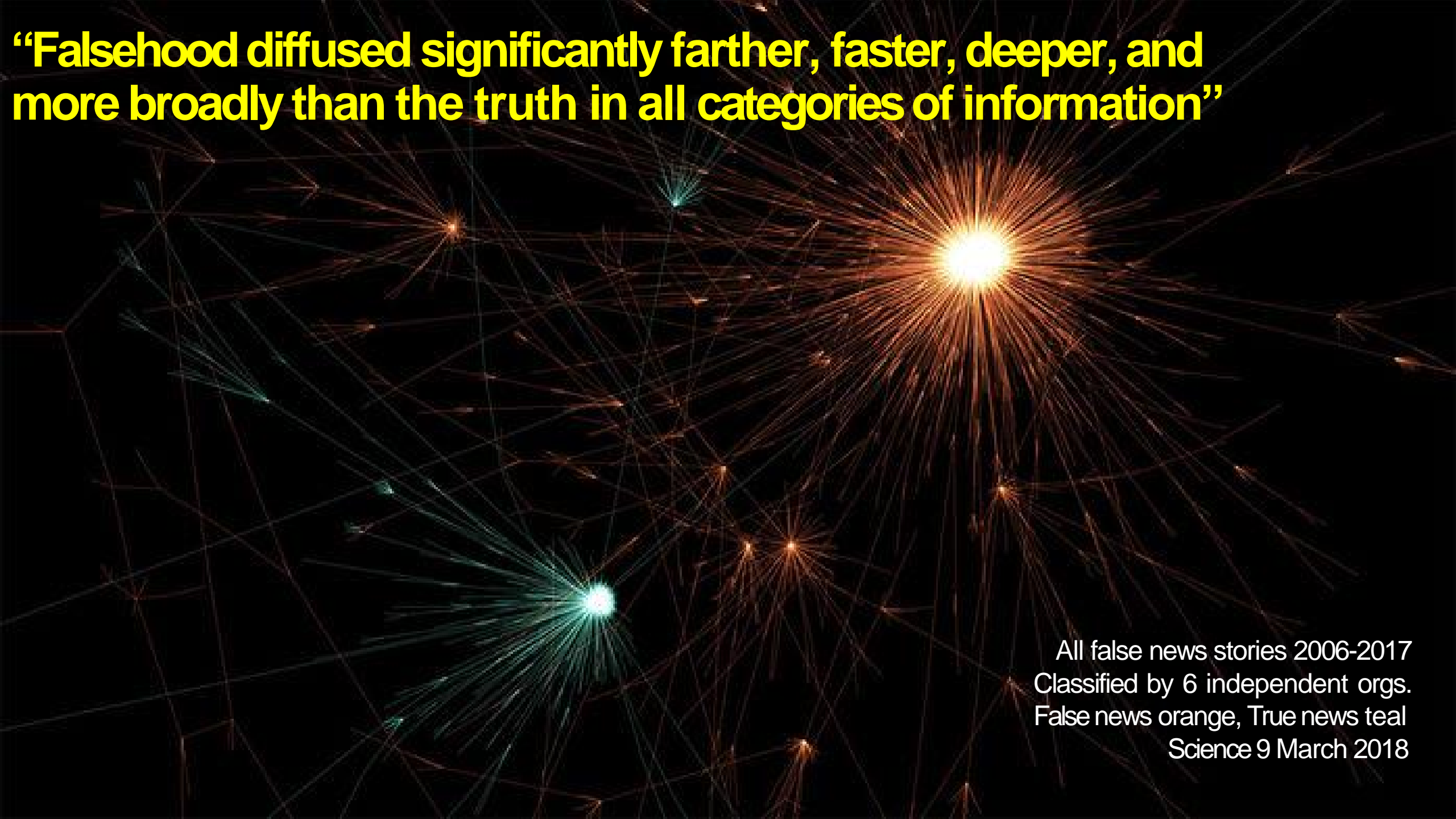
What is Fake News

- ? **Falsehood or Misinformation**
- ? **Propaganda**
- ? **Parody**
- ? **Intentional Controversy**
- ? **Unintentional Errors**



“We will take back control of roughly £350m per week”
Brexiteer Boris Johnson. The true figure, £250m, is dwarfed
by the 1.3% decline in GDP in 2018 (and 3.3% in 2020).

“Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information”



All false news stories 2006-2017
Classified by 6 independent orgs.
False news orange, True news teal
Science 9 March 2018

Fake News is Not New



In the “Court of Star Chamber” they had no defense.
It only made Kingardner/for Ghedrown/Cozef and himself.

Why is Fake News a Problem?

- **Decision Error from False Information**
 - Miscast votes, disenfranchised voters
 - Belief in, propagation of revisionist history
 - Wrong foods, medicines, prophylactics
- **Balkanized Society:**
 - Pro vs Anti Abortion
 - Pro vs Anti Brexit
 - Pro vs Anti Gun Control...
 - Pro vs Anti Immigration
 - Pro vs Anti Vaccination
- **3rd Party Harm**
 - Hoax cow killing posted to WhatsApp led to lynching in India
 - Hoax Clinton child slavery ring in New Jersey led to man fire three shots in pizzeria
 - Russian hacker published fake news on state run Qatar News site, inflaming diplomatic row with Saudi Arabia

Why is Fake News a *Hard* Problem?

- **Decision Error from False Information**
 - Miscast votes, disenfranchised voters
 - Belief in, propagation of revisionist history
 - Wrong foods, medicines, prophylactics
- **Balkanized Society:**
 - Pro vs Anti Abortion
 - Pro vs Anti Brexit
 - Pro vs Anti Gun Control...
 - Pro vs Anti Immigration
 - Pro vs Anti Vaccination
- **3rd Party Harm**
 - Hoax cow killing posted to WhatsApp led to lynching in India
 - Hoax Clinton child slavery ring in New Jersey led to man fire three shots in pizzeria
 - Russian hacker published fake news on state run Qatar News site, inflaming diplomatic row with Saudi Arabia
- **Veracity:** Certifying truth is very hard (e.g. Bush v Gore)
- **Liability:** Unlike BBC or NYT, FB and Twitter do not make their news
- **Secrecy:** Targeting whole populations *privately* is possible
- **Velocity:** Almost frictionless dissemination without verification
- **Tenacity:** Truth does not always replace fiction. Irreversibility.

**“A new system of
safeguards is needed.”**

Science of Fake News – Lazer, Baum, Benkler, Berinsky, Greenhill,
Menezer, Metzger, Nyhan, Pennycook, Rothschild, Schudson, Sloman,
Sunstein, Thorson, Watts, Zittrain

Motives: *Why* Do People Share Fake News?

1. Macedonian teenagers do it for fun & profit
2. Vladimir Putin does it as sovereign interference to influence votes & divide the West
3. Citizen Ideologues do it assuming the “ends justify the means”

Solutions 1.0

1. Fun & Profit

2. Sovereign Interference

3. Citizen Ideologues

1. Use economics to change incentives. Apply penalties, warrantees, and escrow to ad driven revenue

2. Small fines no match for state actors. Identify & ban. Apply state level sanction. Putin has no right to vote in UK, FR, US.

3. Challenge: Citizens have a right to their views. Ideologues don't respond well to economic sanction

Solutions (in use) 1.5

- Crowdsourcing Truth / Algorithmic Truth – Idea is impartiality
- Tagging & Product Labeling – Idea is to inform consumers like salt, sugar, fat; debias msg, knowing source
- Ban Individuals – To remove problem
- Educate Consumers – To reduce influence
- Truth “Chasers” & Improved discovery – To undo influence
- Demote stories in news feed

Solutions (in use) 1.5

- Crowdsource Truth / Algorithmic Truth – Idea is impartiality
 - Very hard to prove a claim did NOT happen, e.g. Trump rally with better pic
 - Moody's credit crisis problem (not incentive compatible if FB pays)
 - NYT State capital problem (NYC vs Albany)
- Tagging & Product Labeling – Idea is to inform consumers like salt, sugar, fat; debias msg, knowing source
- Ban Individuals – To remove problem
- Educate Consumers – To reduce influence
- Truth "Chasers" & Improved discovery – To undue influence
- Demote stories in news feed

Solutions (in use) 1.5

- Crowdsource Truth / Algorithmic Truth – Idea is impartiality
 - Very hard to prove a claim did NOT happen, e.g. Trump rally with better pic
 - Moody's credit crisis problem (not incentive compatible if FBpays)
 - NYT State capital problem (NYC vs Albany)
- Tagging & Product Labeling – Idea is to inform consumers like salt, sugar, fat; debias msg, knowing source
 - Ideologues don't care; they just hate Hilary
 - Recursive. Just discredit the labeler. Constant attacks on "biased" media
 - Content labels do not provide means to re-contact affected audience ("white male coal miner in Appalachia" ≠ Ch 5 News)
- Ban Individuals – To remove problem
- Educate Consumers – To reduce influence
- Truth "Chasers" & Improved discovery – To undue influence
- Demote stories in news feed

Solutions (in use) 1.5

- Crowdsourced Truth / Algorithmic Truth – Idea is impartiality
 - Very hard to prove a claim did NOT happen, e.g. Trump rally with better pic
 - Moody's credit crisis problem (not incentive compatible if FB pays)
 - NYT State capital problem (NYC vs Albany)
- Tagging & Product Labeling – Idea is to inform consumers like salt, sugar, fat; debias msg, knowing source
 - Ideologues don't care; they just hate Hilary
 - Recursive. Just discredit the labeler. Constant attacks on "biased" media
 - Content labels do not provide means to re-contact affected audience ("white male coal miner in Appalachia" ≠ Ch 5 News)
- Ban Individuals – To remove problem
 - Arms race, just return under new ID
 - Citizens have right to voice their opinion
- Educate Consumers – To reduce influence

- Truth "Chasers" & Improved discovery – To undue influence

- Demote stories in news feed

Solutions (in use) 1.5

- Crowdsourced Truth / Algorithmic Truth – Idea is impartiality
 - Very hard to prove a claim did NOT happen, e.g. Trump rally with better pic
 - Moody's credit crisis problem (not incentive compatible if FB pays)
 - NYT State capital problem (NYC vs Albany)
- Tagging & Product Labeling – Idea is to inform consumers like salt, sugar, fat; debias msg, knowing source
 - Ideologues don't care; they just hate Hilary
 - Recursive. Just discredit the labeler. Constant attacks on "biased" media
 - Content labels do not provide means to re-contact affected audience ("white male coal miner in Appalachia" ≠ Ch 5 News)
- Ban Individuals – To remove problem
 - Arms race, just return under new ID
 - Citizens have right to voice their opinion
- Educate Consumers – To reduce influence
 - Context on context creates information overload
 - Confirmation bias, selective hearing
 - Many distrust those who would "educate" them
- Truth "Chasers" & Improved discovery – To undue influence
- Demote stories in news feed

Solutions (in use) 1.5

- Crowdsourc Truth / Algorithmic Truth – Idea is impartiality
 - Very hard to prove a claim did NOT happen, e.g. Trump rally with better pic
 - Moody's credit crisis problem (not incentive compatible if FBpays)
 - NYT State capital problem (NYCvs Albany)
- Tagging & Product Labeling – Idea is to inform consumers like salt, sugar, fat; debias msg, knowing source
 - Ideologues don't care; they just hate Hilary
 - Recursive. Just discredit the labeler. Constant attacks on "biased" media
 - Content labels do not provide means to re-contact affected audience ("white male coal miner in Appalachia" ≠ Ch 5 News)
- Ban Individuals – To remove problem
 - Arms race, just return under new ID
 - Citizens have right to voice their opinion
- Educate Consumers – To reduce influence
 - Context on context creates information overload
 - Confirmation bias, selective hearing
 - Many distrust those who would "educate" them
- Truth "Chasers" & Improved discovery – To undue influence
 - Evidence suggests these have almost no effect (BNyhan)
- Demote stories in news feed

Solutions (in use) 1.5

- Crowdsourced Truth / Algorithmic Truth – Idea is impartiality
 - Very hard to prove a claim did NOT happen, e.g. Trump rally with better pic
 - Moody's credit crisis problem (not incentive compatible if FB pays)
 - NYT State capital problem (NYC vs Albany)
- Tagging & Product Labeling – Idea is to inform consumers like salt, sugar, fat; debias msg, knowing source
 - Ideologues don't care; they just hate Hilary
 - Recursive. Just discredit the labeler. Constant attacks on "biased" media
 - Content labels do not provide means to re-contact affected audience ("white male coal miner in Appalachia" ≠ Ch 5 News)
- Ban Individuals – To remove problem
 - Arms race, just return under new ID
 - Citizens have right to voice their opinion
- Educate Consumers – To reduce influence
 - Context on context creates information overload
 - Confirmation bias, selective hearing
 - Many distrust those who would "educate" them
- Truth "Chasers" & Improved discovery – To undue influence
 - Evidence suggests these have almost no effect (BNyhan)
- Demote stories in news feed
 - Like spam, can simply motivate variation to seek to get through

**Which of these effectively
change incentives to produce
disinformation?**

Solutions (in use) 1.5

- Apply incentives & penalties – to curb ad driven disinformation
- Crowdsourced Truth / Algorithmic Truth – Idea is impartiality
- Tagging & Product Labeling – Idea is to inform consumers like salt, sugar, fat; debias msg, knowing source
- Ban Individuals – To remove problem
- Educate Consumers – To reduce influence
- Truth “Chasers” & Improved discovery – To undue influence
- Demoting in News Feed – To undue influence

Add Friction: Downsample & Delay

This method puts determination of original truth back on source. Knowing they will lose influence, they will choose truth more often.

Solution I :

Add Friction to Dissemination of False News

Rule 1: If reputation is sterling, number of followers and speed of news delivery are unlimited , but...

Rule 2: If caught propagating false news, either as source or sharer, followers are limited and dissemination is delayed

Expost verification is far easier than ex ante. Places burden back on source to be truthful.

Trolls, who are banned and return, cannot build followers using fake news to build engagement.

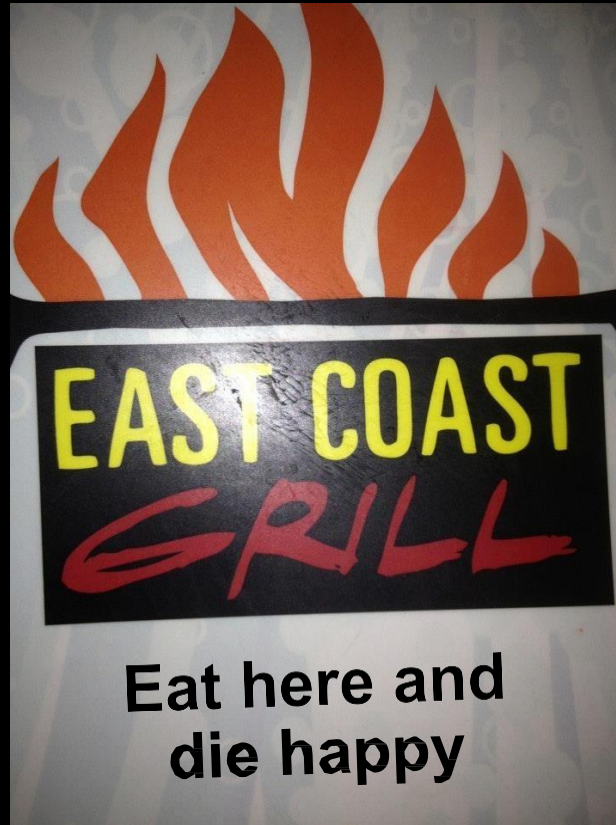
Ideologues need access to disseminate views. Reputations quickly lost and slowly gained suggest even ideologues behave better

Pressing on concepts...

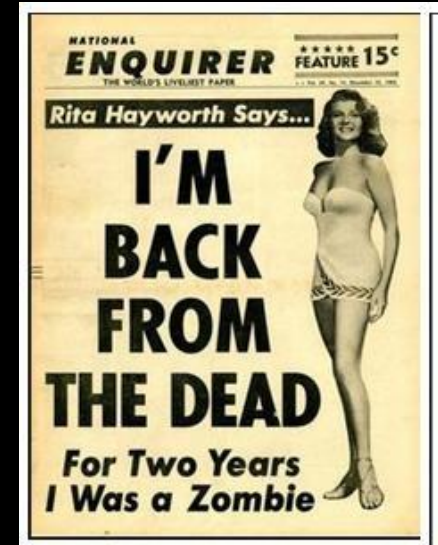
Fact vs Fiction: Truth Alone is *Not* the Issue



Pluto is a planet, is not, is, is not...



Eat here and die happy



Rita Hayworth alive again!

The truth of Pluto's status, claims of longevity, or an actor's vivacity are all scientifically verifiable facts. They're just not important.

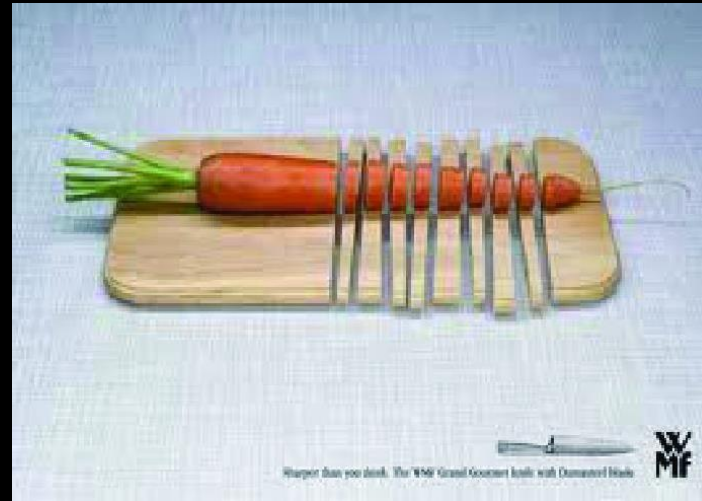
Fact vs Fiction: Truth Alone is *Not* the Issue



World's hottest
pepper



World's biggest
cheese taste



World's sharpest
smoothest shave



World's most
comfortable chair

Exaggerated truth has long been used for rhetorical or mnemonic effect.

Fact vs Fiction: Truth Alone is *Not* the Issue

Selective



Truth

Trump supporter argues Elkhart Indiana unemployment 20% under Obama and 3% under Trump ...but omits fact it already fell to 4% under Obama

True information can be presented so as to tell opposing stories

Fact vs Fiction: Truth Alone is *Not* the Issue

True

True



Russian troll accounts flood twitter after Parkland, when black men were shot by white police and during black history month to balkanize voters.

Much false information doesn't matter. Much true information does

**If truth isn't the issue, maybe
it's harm...**

Good vs Bad: Harm Alone is *Not* the Issue



Vitaly Borker elevated page rank by cheating and threatening customers

Google fixed algorithm



Apple exercises bouncer's rights to exclude porn, hate speech & viruses



Uber protects riders against "fake routes" and drivers against people getting sick in their cars



Airbnb offers \$1M insurance to protect against loss
...backup rooms to protect against misrepresentation

Platforms try to protect their users against mistreatment.

Good vs Bad: Harm Alone is *Not* the Issue



Alibaba developed AliPay as an escrow service to protect users against fraud



Viddy spammed users' friends to adopt its video service

Facebook terminated API access



eBay punishes buyers who don't pay sellers, who don't ship, and merchants who lie about their products

Platforms try to protect their users against mistreatment.
If harm isn't the issue, what fails in case of fake news?

Insight from Fake Goods

Truth



\$3,000 Bottega Veneta

Fiction



\$50 Bottega Veneta

With certificate of authenticity

The seller does not complain. The buyer does not complain. The platform does not complain
Bottega Veneta, off platform, complains.

Fake News

Truth



Clinton Foundation convenes businesses and NGOs to improve global health

Fiction



Accused of sponsoring occult religions, sex trafficking, paying for Chelsea wedding

The creator does not complain. The sharer does not complain. FB did not complain
Hillary Clinton, off platform, complains.

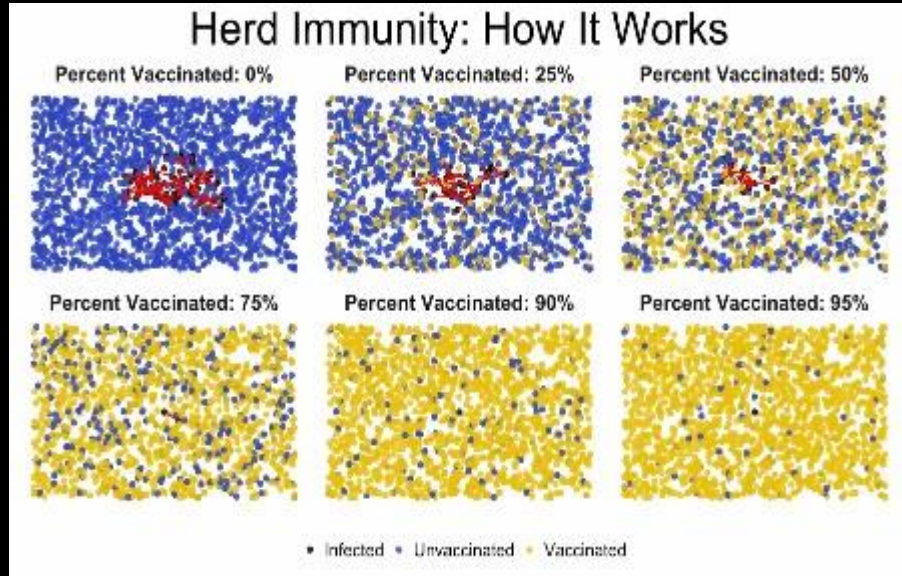
**Real Issue is externalities, harm
that occurs *off* platform**

A photograph of Mark Zuckerberg speaking at a hearing. He is wearing a dark blue suit, a white shirt, and a blue patterned tie. He is looking slightly to his right. The background is blurred, showing other people in suits.

**“We did not take a broad
enough view of our
responsibility...”**

Mark Zuckerberg
Joint Commerce & Judiciary Committee
April 10, 2018

Real Issue is Harm Occurring *Off* Platform



Anti Vaxxer campaigns promote disease outbreak



Precedent Trump



Shots fired over pizzagate

While platforms correct market failures that occur on platform, they have not demonstrated willingness to correct failures that occur off platform

Solution II :

Join Decision Rights & Knowledge of Harm

Option 1: Pull outside information into platform

- X Creates a powerful central platform
- X Does not ensure incentive alignment
- X Infinite variety of externalities suggests many missed

Option 2: Push inside information out where externalities occur

- ✓ Creates competition, decentralization
- ✓ Solves secrecy problem. Clinton can address messages whispered by Russians
- ✓ Informed parties can act. Affected parties can target source

**Objection! Our data, our algorithms
and our ad models are trade secrets.**

Response: Trade secret is not absolute. We allow trade secrets in order to prevent misappropriation and unjust enrichment. The incentive to innovate not only creates value per se but also yields a positive externality at the point 3rd parties uncover the secret.

Externalities of this secret are negative not positive. It is producing more harm than good.

Solution IIb : Transparency

Rule 3: Void the offending trade secret - the greater the potential for negative externalities, the sooner such secrets are void.

Rule 4: Craft disclosure narrowly, i.e. only just broad enough to address the externality.

Consider users' personal data, friend lists, and posts. Consider ad buyers' identity, content, targets, and spend.

Non-political ads for biggest cheese taste, hottest pepper, and closest shave have almost no externalities.

Political ads and policy debates involve externalities. *Full transaction disclosure* of ad buyer identities, targets, content, and spend does not compromise friend lists or prior posts

**Knowing the law, Facebook et. al. can *choose* whether
to accept ads.**

Knowing the law, Facebook et. al. can *choose* whether to accept ads.

If an ad category holds the prospect of harmful externalities (exogenously determined), FB can choose to make money knowing it must disclose its trade secrets, else it can forgo the activity that causes harm.

Knowing the law, Facebook et. al. can *choose* whether to accept ads.

If an ad category holds the prospect of harmful externalities (exogenously determined), FB can choose to make money knowing it must disclose its trade secrets, else it can forgo the activity that causes harm.

The firm internalizes its damage, and the estimate of business value relative to social cost moves toward socially optimal

Summary

- For reasons of veracity, liability, tenacity, secrecy, velocity, fake news is a hard problem
- Motives include economics, state interference, and ideology
- Fake news is less about veracity than about harm
- Fake news is less about harm on-platform than harm off-platform
- Standard solutions – taxing bad behavior, crowdsourcing truth, banning trolls, labeling content, demoting, and educating users – rarely address underlying issues:
 1. Sources & sharers are not motivated to change behavior
 2. knowledge of the harm is not paired with decision rights in the transaction.

Summary

- The solutions are to
 1. Publicly apply friction to communication of agents who share false information, on both time and contacts
 2. Void the offending trade secret that gives rise to a negative externality. The greater the risk, the more immediate the disclosure. Disclose just broadly enough to pair knowledge of the transaction with knowledge of the harm.